Thesis for the degree of Licentiate of Philosophy

New perspectives on mesospheric wave dynamics and oxygen photochemistry

Anqi Li

Department of Space, Earth and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2020

New perspectives on mesospheric wave dynamics and oxygen photochemistry
Anqi Li

Licentiatavhandlingar vid Chalmers tekniska högskola

Department of Space, Earth and Environment
Chalmers University of Technology
SE–412 96 Göteborg, Sweden
Telephone + 46 (0) 31 – 772 1000

Cover: An orbit of the retrieved daytime volume emission rate of $O_2(a^1\Delta_g)$ measured by OSIRIS, same as the upper panel of Fig.6.1

Typeset by the author using LaTeX.

Printed by Chalmers digitaltryck
Göteborg, Sweden 2020

Abstract

The mesosphere is the region of the atmosphere between 50 km to 100 km, where both dynamical and photochemical aspects play important roles for the thermal balance. This thesis focuses on the following three areas for mesospheric studies: wave dynamics, oxygen photochemistry and retrieval using the optimal estimation method.

Atmospheric gravity waves are internal disturbances in the medium that propagate horizontally and vertically. Based on linear wave theory, this thesis attempts to enhance our understanding of the relationships between the wave characteristics, the mean flow and the sources. We try to emphasise the frequency change due to the Doppler effect in several reference frames. This thesis proposes a consistent framework for deriving those wave parameters that cannot be obtained from a single type of instrument due to their particular observational geometry. Finally, a plausible interpretation of a readily available ground-based lidar observation is given as an example.

Oxygen photochemistry is another important aspect in this thesis. The underlying chemical reactions are affected by disturbances in the local temperature and density, which in turn changes the distribution of the excited oxygen species. In this work, a photochemical model has been implemented, which describes most of the important processes such as O_3 photolysis that are related to the production and loss of $O(^1D)$, $O_2(b^1\Sigma_g^+)$ and $O_2(a^1\Delta_g)$.

The observation of airglow emissions provides an opportunity to explore the chemical composition and wave dynamics in the upper mesosphere. The Odin satellite has been routinely measuring $O_2(a^1\Delta_g)$ airglow emissions since 2001. In this thesis, data collected by OSIRIS are explored. Inversions are carried out in order to retrieve the volume emission rate of $O_2(a^1\Delta_g)$ as well as the mesospheric ozone density. The resulting ozone profiles are shown to be consistent with other independent ozone datasets collected by instruments aboard the same spacecraft as well as ACE-FTS and MIPAS, despite intrinsically different measurement principles. The overall good agreement between them illustrates the good performance of the retrieval technique. Furthermore, these investigations serve well as a preparatory activity for the upcoming satellite mission MATS, set for launch later this year.

Keywords: Satellite limb observation, gravity wave, oxygen airglow, mesospheric ozone.

Acknowledgments

Writing the acknowledgement part of a thesis has always been a process of trying to describe the internal feelings in words with my limited vocabulary – it is not easy. Thank, in my opinion, is a word that is beyond my sincere gratefulness towards all people who assist on my, occasionally bumpy, journey of pursuing a PhD education. But I will try, as much as I can.

The main supervisor and co-supervisors are, of course, the first ones who I want to say thank you to. Donal, Kristell and Ole Martin, I am very grateful to have your continuous support in both my academic and daily life. Your patience to accept, to tolerate and to resolve my countless questions has always been the greatest relief and source to let me grow. Your in-depth knowledge of the scientific fields and encouragements pave the way for my development both as a researcher and as a person. Thank you, Sensei!

I would also like to send my gratitude over the ocean to Chris, Adam and Doug in Saskatoon for your valuable inputs and continuous interests in my work. Your comprehensive guidance and enthusiasm on the OSIRIS project make it pleasurable to collaborate with you. Also, many thanks go to Andreas who has hosted me at DLR Oberpfaffenhofen and provided active discussions about gravity wave science. I appreciate Linda and Jörg in Stockholm for including me in the MATS team and for your interests and trusts in my wavy work. To my PhD colleagues in the department and Paulina, thank you for sharing your coding skills, your critical thoughts, your openness, laughter, concerns, administrative supports...which make everything easier and fun.

Finally, I would like to deliver my deepest appreciation to my family for all the care and encouragement over the long period of more than a decade of exotic life. It would have otherwise been a complete different version of me. And to my dear Alex who has the greatest patience, thanks for accompanying me on the journey of life side by side.

Anqi Li
Göteborg, March 2020

List of Publications

This thesis is based on the following appended papers:

Paper 1. Anqi Li, Chris Roth, Kristell Pérot, Ole Martin Christensen, Adam Bourassa, Doug Degenstein and Donal Murtagh. *Retrieval of daytime mesospheric ozone using OSIRIS observation of $O_2(a^1\Delta_g)$ emission* Under review (2020).

Paper 2. Anqi Li, Ole Martin Christensen, Andreas Dörnbrack, Patrick Espy, Alexander Kozlovsky, Mark Lester, Robert Reichert and Donal Murtagh. *Perspectives on interpretation of a gravity wave event recorded by ground-based lidar* Manuscript in preparation.

Other relevant publications co-authored by Anqi Li:

Jörg Gumbel, Linda Megner, Ole Martin Christensen, Nickolay Ivchenko, Donal P. Murtagh, Seunghyuk Chang, Joachim Dillner, Terese Ekebrand, Gabriel Giono, Arvid Hammar, Jonas Hedin, Bodil Karlsson, Mikael Krus, **Anqi Li**, Steven McCallion, Georgi Olentšenko, Soojong Pak, Woojin Park, Jordan Rouse, Jacek Stegman, Georg Witt. *The MATS satellite mission-Gravity wave studies by Mesospheric Airglow/Aerosol Tomography and Spectroscopy* Atmospheric Chemistry and Physics 2020 vol: 20 (1) pp: 431-455.

Contents

Abstract iii

Acknowledgments v

List of Publications vii

I Introductory chapters 1

1 Welcome to the mesosphere 3
1.1 Why is the mesosphere so interesting? . . . 3
1.2 Current observations with Odin . . . 5
1.3 Observations with MATS in the future . . . 5

2 Mesospheric wave dynamics 9
2.1 Introduction . . . 9
2.2 Fundamental linear wave theory . . . 10
2.2.1 The effects of mean flow . . . 12
2.2.2 Gravity wave sources . . . 16
2.3 Observations and interpretation . . . 18
2.3.1 Imager observations . . . 19
2.3.2 Ground-based vertical profilers . . . 21

3 Mesospheric photochemistry 23
3.1 Introduction . . . 23
3.2 Fundamental chemical kinetics . . . 25
3.2.1 The steady state assumption . . . 25
3.2.2 Photochemical processes . . . 27
3.3 Oxygen airglow photochemistry . . . 28

4 The inversion problem 35
4.1 Introduction . . . 35
4.2 Theory of optimal estimation in Bayesian philosophy . . . 36
4.2.1 Linear optimal estimation . . . 37
4.2.2 Non-linear optimal estimation . . . 39
4.3 Practical implementation . . . 41

5 Summary of appended publications **45**

6 Outlook **47**
6.1 Expansion of the retrievals 47
6.2 Mesospheric science studies 49
6.3 The future mission MATS 51

Bibliography **53**

II Appended papers **59**

1 Retrieval of daytime mesospheric ozone using OSIRIS observation of O_2 ($a^1\Delta_g$) emission **61**

2 Perspectives on interpretation of a gravity wave event recorded by ground-based lidar **93**

Part I

Introductory chapters

Chapter 1

Welcome to the mesosphere

Earth's atmosphere is divided into different layers based on their characteristics. One of the most common divisions of the atmosphere is based on its vertical temperature gradient which is either positive or negative, as shown in Fig. 1.1. The mesosphere[1] is one of the 'spheres' among others, where the temperature gradient is negative and it ranges approximately from 50 km to 100 km altitude. Other 'spheres' such as the troposphere, stratosphere and thermosphere are not studied explicitly in this thesis, although some interesting phenomena to be discussed reach slightly above the boundary of mesosphere and together the region is commonly known as the MLT region (i.e., the mesosphere and lower thermosphere).

The examination of mesospheric processes is a **multi-disciplinary study**. Thermodynamical and dynamical aspects are commonly discussed separately in the domain of meteorology. In the subject of aeronomy, chemical and photochemical aspects are often distinguished. However, in the mesospheric region which we address here, these subjects interact with each other and the coupling effects of them shall be considered carefully. In addition, while we are handling data collected from remote sensing techniques and attempt to convert from the measured quantity to the desired quantity, other disciplines shall also be addressed that are the estimation theory and signal processing. This thesis focuses on the following three main areas: wave dynamics, oxygen photochemistry and retrieval using the optimal estimation method.

1.1 Why is the mesosphere so interesting?

The top of the mesosphere, the mesopause, is often not clearly defined because the exact altitude varies with latitude and seasons, but it can be considered as a region between 80 km and 100 km. Here, the coldest temperatures of Earth's atmosphere can be found. More precisely, the coldest place is located at the summer pole, rather than winter pole where normally one would expect it to be. In turn, the extreme cold temperatures at the summer mesopause help to create noctilucent clouds (NLCs), despite the fact that the air pressure humidity is so low there. So why would it be so cold at the summer pole even though the solar heating rate is larger there than in

[1] from Greek 'mesos', middle

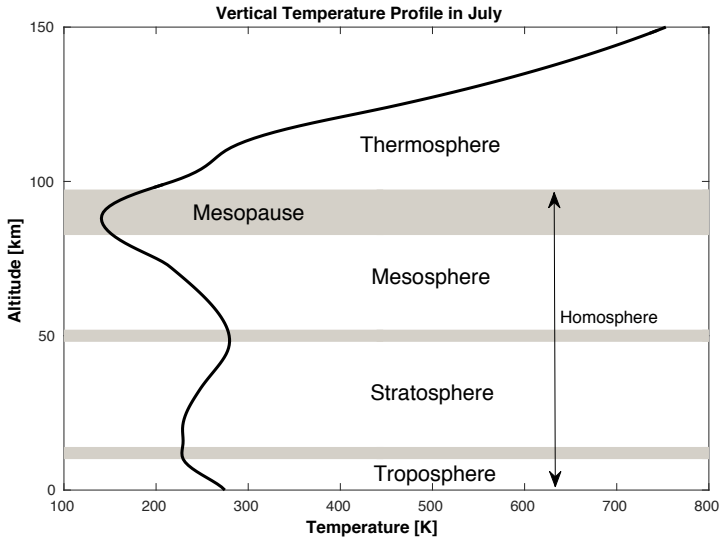

Figure 1.1: Divisions of the atmosphere based on vertical temperature gradient or composition. Vertical temperature profile is obtained from MSISE90 model in July at latitude of 80 °S

the winter hemisphere? A simple explanation is that the extreme cold temperatures are due to the effects to the net upward rising motion and the associated adiabatic cooling process. This is one of the examples showing that the dynamical aspect modifies the thermal balance in the atmosphere. In addition to the adiabatic cooling and heating mechanisms, processes such as the transport of thermal energy by eddy diffusion caused by gravity waves breaking, heating due to chemical reactions (e.g., exothermic heating) as well as radiative cooling by carbon dioxide are all playing their crucial roles in the energy budget of this high altitude region.

Another distinct feature in the mesosphere, besides the extreme temperature at the top, is the existence of 'glowing gases' as known as airglows that is a result from the active chemical interactions with the solar light in the ultraviolet and visible spectrum. Airglow is similar to aurora which both appear in the same altitude region and are linked to the chemical activity of a similar group of gases. In contrast to aurora, airglow corresponds to the light emitted by gases produced by photochemical reactions and distributed around the globe. The uneven structure of the airglow reflects local temperature and chemical species distribution and thus the observation of it help us to understand the internal variability produced by atmospheric waves.

The mesosphere is a very difficult altitude region to probe by in-situ measurements. The air is too thin at that altitude and cannot provide enough buoyancy to lift an aircraft or even balloon (commonly used to study the troposphere and stratosphere). On the other hand, the atmospheric friction is too large for an orbital spacecraft to fly around. The only feasible way to access the altitude region is by sending sounding rockets which allows to take measurements for a few minutes per mission. Thus, remote sensing plays an important role in the study of the mesosphere. The Swedish National Space Agency (SNSA) is involved in two satellite missions that support the

scientific investigation of the mesosphere, namely Odin and MATS. The following sections provide a brief summary of the features of these two satellites.

1.2 Current observations with Odin

Since 2001 the Odin satellite is orbiting Earth at ca. 600 km altitude, around 15 times per day and is fully active to date (D. Murtagh et al. 2002). The main scientific objectives of Odin include middle atmospheric ozone and NLC sciences, as well as the coupling between the upper and lower atmosphere. The two main payloads on Odin are SMR (Submillimeterwave Radiometer, measures 486-580 GHz and at 119 GHz) and OSIRIS (Optical Spectrograph and Infrared Imaging System, measure 274-810 nm and 1255-1275 nm, 1510-1550 nm, respectively). Both of them measure the concentration of various species closely related to the ozone chemistry such as NOx, CO, H_2O, ClO, N_2O, HNO_3 by observing Earth's limb. Furthermore, OSIRIS consists of two optically independent instruments: the optical spectrograph (hereafter OS) and the infrared imager (hereafter IRI). IRI has three vertical channels. Two of them measure the oxygen infrared atmospheric (0-0) band (IRA-band) emissions centred at 1.27 µm and one of them measures the OH Meinel (3-1) band emission centred at 1.53 µm. Two example orbits of the measured limb radiance by channel 1 and 3 are shown in Fig 1.2.

1.3 Observations with MATS in the future

Mesospheric Airglow/Aerosol Tomography and Spectroscopy (MATS) is a future satellite mission (Gumbel et al. 2020). The research satellite is scheduled to be launched in the end of 2020 into a 600 km sun-synchronous orbit. The main scientific objective of MATS is to determine the global distribution of gravity waves and other dynamical structures in the MLT region. Two instruments, a limb imager and a nadir imager will be on board. Observations will primarily be done by the limb imager at six spectral channels, with two in the ultra-violet (UV) spectral region centred at 270 nm and 305 nm and four in the infrared spectral region centred at around 762 nm. These spectral channels are designed to target NLCs and oxygen airglow in order to study the dynamical structures, as the atmospheric waves can alter the homogeneity of NLCs and airglow in A-band emission. Figure 1.3 shows a virtual limb image of MATS simulated by a model that couples the oxygen airglow photochemistry and gravity wave (Li 2017).

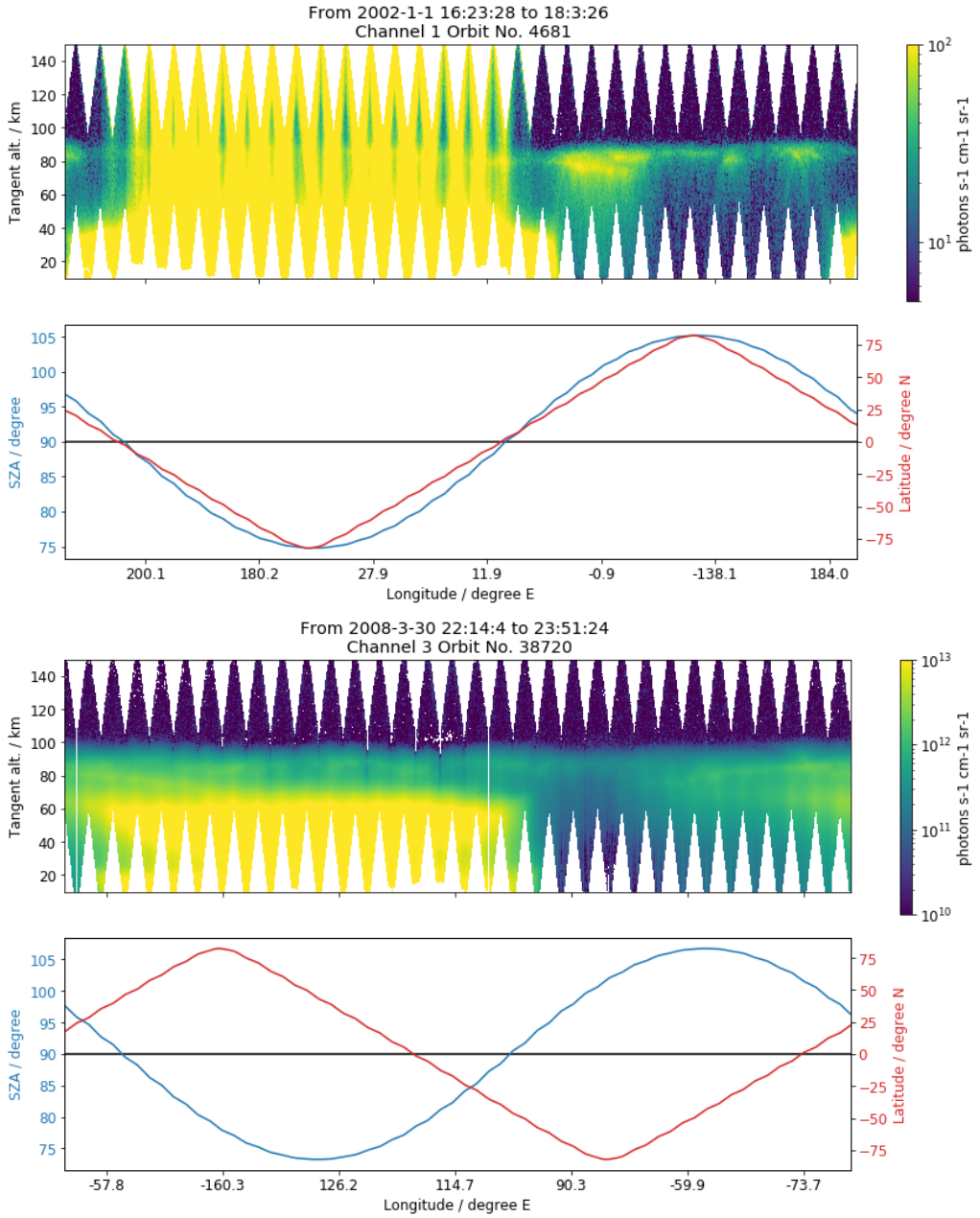

Figure 1.2: Two example orbits of the limb radiance measured by Odin/OSIRIS infrared imager channel 1 (upper) and 3 (lower), spectral band centred at 1.27 µm and 1.53 µm, respectively. The red curves indicate the latitude coordinate at tangent point, whereas the longitude coordinate is shown at the horizontal axes. The blue curves indicate the local solar zenith angle (SZA) at tangent point. The day part of the orbit corresponds to a SZA lower than 90°, and vice versa.

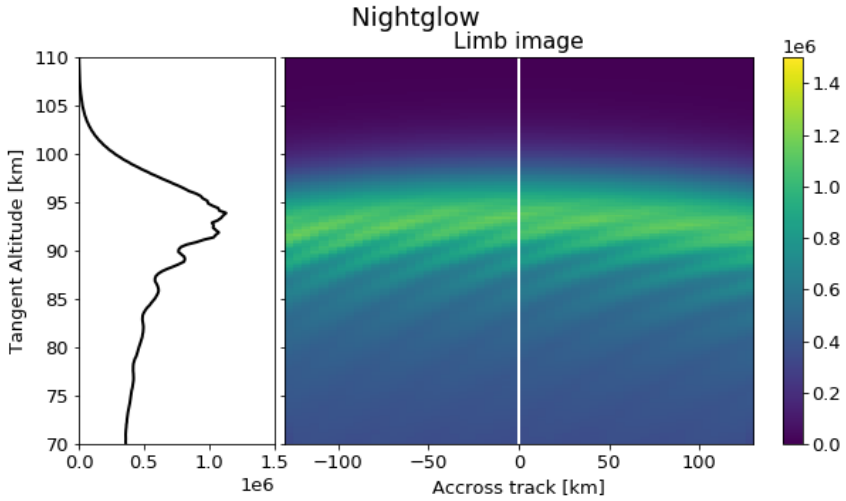

Figure 1.3: A virtual limb image taken from the MATS satellite. The simulated limb image is based on the viewing geometry of MATS measuring an atmospheric volume that has a mono-chromatic wave presented under a background wind shear. The left panel shows the vertical profile of the limb radiance that corresponds to the white vertical line indicated on limb image to the right. Values of the limb radiance in this figure are in an arbitrary unit. The model couples the $O_2(b^1\Sigma_g^+)$ photochemistry and gravity wave, and is described in Li (2017).

Chapter 2

Mesospheric wave dynamics

2.1 Introduction

Atmospheric waves are dynamical features caused by, among other things, the forcing of air parcels over obstacles such as mountains, frontal systems, etc. The key condition for maintaining these waves in the atmosphere is achieved by a restoring force to balance the acceleration caused by the disturbances. The restoring force can be the result of buoyancy in a stably stratified environment. However, in most cases in the middle atmosphere, the Coriolis force is also part of the restoring force especially if the horizontal wavelength of the waves is larger than few hundred kilometres. These small to medium scale waves (typically described by their horizontal scales in few dozens to thousands kilometres) are commonly referred to as *inertia-gravity waves*, hereafter referred to simply as *gravity waves*, and will be the focus for the wave discussion part of this thesis. Larger scale *inertial waves* such as planetary-scale waves, tides etc. will not be addressed further in this thesis.

Once gravity waves are excited near the surface of Earth, they may propagate upwards and act as a coupling agent between the lower and higher atmosphere. Their amplitudes grow substantially in the mesosphere due to the decreasing air density (i.e., conservation of kinetic energy). When the amplitudes are so large that the restoring force can not be maintained, these wave eventually break and deposit their energy to the mean flow. The influence on the mean atmospheric state due to the breaking of these waves becomes important at higher altitudes. In fact, the global circulation in the mesosphere is rather dynamically than thermally driven (Holton 1982; Karlsson and Shepherd 2018). However, large uncertainty in the characterisation of gravity waves on the global scale limits the ability of general circulation models to reproduce phenomena in the real atmosphere (McLandress 1998).

In this chapter, we will start by discussing the basic linear theory of gravity wave characterisation. The theory can then be applied to interpret the observations based on what wave parameters have been measured and further derive the other unknown wave parameters.

2.2 Fundamental linear wave theory

Let us begin with a generic description of a sinusoidal wave form to represent a displacement field ξ

$$\xi = A\cos\phi \tag{2.1}$$

or in exponential notation

$$\xi = \Re(Ae^{i\phi}), \tag{2.2}$$

where $\Re$ denotes the real part of a complex number, A the amplitude of the wave and ϕ the phase angle which is a function of space and time

$$\phi(\vec{r}, t) = \vec{k}\cdot\vec{r} - \Omega t, \tag{2.3}$$

where $\vec{r}$ is the position vector, t the time and $\vec{k}$ the wave vector, which points in the direction of the travelling wave and Ω is the wave frequency. Both $\vec{k}$ and Ω are fundamental properties of a wave. If we look at the space in the Cartesian coordinate system, i.e., $\vec{r} = (x, y, z)$, the wave vector can be decomposed into three components along each direction, i.e.,

$$\vec{k} = (k_x, k_y, k_z). \tag{2.4}$$

Conventionally, we often use x and y to denote the two directions on the horizontal plane and z to denote the vertical direction. k_x and k_y sometimes can be composed into the 'total' horizontal component, k_h, if we focus on a 2 dimensional problem. Each component is termed the wavenumber which can be thought as 2π times the number of wave oscillations per unit length, or wavelength per unit radian. Likewise, the wave frequency is 2π times the number of wave oscillations per unit time or radians per second. In other words,

$$(k_x, k_y, k_z, \Omega) = (\frac{2\pi}{\lambda_x}, \frac{2\pi}{\lambda_y}, \frac{2\pi}{\lambda_z}, \frac{2\pi}{\tau}), \tag{2.5}$$

where λ_x, λ_y and λ_z are wavelengths in each direction and τ the wave period. The phase speed of the wave, c, the speed at which constant phase moves in the direction of the travelling wave, can also be seen (and is measurable) along each direction, i.e.,

$$(c_x, c_y, c_z) = (\frac{\Omega}{k_x}, \frac{\Omega}{k_y}, \frac{\Omega}{k_z}) = (\frac{\lambda_x}{\tau}, \frac{\lambda_y}{\tau}, \frac{\lambda_z}{\tau}). \tag{2.6}$$

However, it is interesting to note that the phase speed of a wave is not a vector and that the total phase speed follows the following relation

$$\frac{1}{c^2} = \frac{1}{c_x^2} + \frac{1}{c_y^2} + \frac{1}{c_z^2}. \tag{2.7}$$

For a given wave period, long waves travel faster than short waves which leads to the wave *dispersion*. The *dispersion relation* connects the wave's spatial characteristics (wavenumbers) and the wave frequency. For internal gravity waves in the atmosphere, the dispersion relation reads

$$k_z^2 = \frac{(k_x^2 + k_y^2)(N^2 - \Omega^2)}{(\Omega^2 - f^2)} - \frac{1}{4H^2}, \tag{2.8}$$

or alternatively

$$\Omega^2 = N^2 \frac{k_x^2 + k_y^2}{k_x^2 + k_y^2 + k_z^2 + 1/(4H^2)} + f^2 \frac{k_z^2 + 1/(4H^2)}{k_x^2 + k_y^2 + k_z^2 + 1/(4H^2)}, \tag{2.9}$$

where N is the buoyancy frequency, also known as the Brunt-Väisälä frequency, f the Coriolis parameter, and H the scale height. This dispersion relation is derived from the fundamental fluid equations, which takes into account the buoyancy, the gravity, and the Coriolis force, but eliminating the viscosity and the acoustic property of the fluid (i.e., the air). Note that although the acoustic wave is eliminated, the compressibility term related to the atmospheric density gradient, the so-called '4H term', is retained in the expression. The detailed derivation can be found in e.g., Holton (1982), Fritts and Alexander (2003), and Nappo (2002). If the Coriolis parameter, f, is negligible compared to Ω and N, and the compressibility term is small, the dispersion relation can be further simplified to

$$k_z^2 = (k_x^2 + k_y^2)\frac{(N^2 - \Omega^2)}{\Omega^2} \tag{2.10}$$

or alternatively

$$\Omega^2 = N^2 \frac{k_x^2 + k_y^2}{k_x^2 + k_y^2 + k_z^2}. \tag{2.11}$$

The Brunt-Väisälä frequency, N, characterises the background atmosphere properties, specifically the vertical temperature gradient. This frequency is related to the difference between the *atmospheric lapse rate*, γ_a and the *adiabatic lapse rate*, Γ

$$\begin{aligned} N^2 &= \frac{g}{T_a}\left(\frac{g}{c_p} - \frac{\partial T_a}{\partial z}\right) \\ &= \frac{g}{T_a}(\Gamma - \gamma_a), \end{aligned} \tag{2.12}$$

where T_a is the atmospheric temperature, c_p the specific heat capacity at constant pressure and g is the gravitational acceleration. The Brunt-Väisälä frequency represents the maximum frequency for vertically propagating gravity waves, that is when the air parcel is purely vertically displaced. In reality, a propagating gravity wave always has a direction at an angle to the vertical, so the air parcel is displaced at an angle to the vertical. Thus the frequency of the gravity wave is expressed as

$$\Omega = N \cos\beta. \tag{2.13}$$

If one compares 2.13 to Eq. 2.11, one can easily find that β is the angle between the wave vector and the horizontal plane.

Of course, the oscillation of the air parcel that has been mentioned above is under the assumption that N is a real number, which is referred to as a *stably-stratified* environment. If N is an imaginary number, i.e., $\gamma_a > \Gamma$, it is referred to as a *convective instability* in the atmosphere, which causes the air parcel to exhibit unbounded growth in response to the vertical displacement. Such condition may

occur especially when the wave amplitude has grown so large at high altitudes that the local temperature eventually exhibits a gradient greater than the adiabatic lapse rate, i.e., a *superadiabatic lapse rate*. At this point, wave breaking occurs and transfers the wave energy into turbulent kinetic energy. Likewise the amplitude of a water surface wave on a coastline reaches a point where the crest of the wave actually overturns. Around the wave breaking region, the linear wave theory often becomes invalid to describe the dynamical behaviour and thus will not be further discussed in this thesis.

2.2.1 The effects of mean flow

Our discussions so far characterise the behaviour of internal gravity waves within the mean flow, as these waves are essentially the perturbations of the mean. We intuitively describe the airflow above the surface as 'moving air', or more commonly, wind. The so-called 'wind speed' and 'wind direction' are always measured relative to the surface ground. Therefore, the characterisation of gravity waves should also be studied in the frame of reference that we often refer to, i.e., the ground. Classically, we tend to think of how the background wind affects the behaviour of internal gravity waves. Though, these waves can not solely exist without the flow field and are bonded together with the background flow regardless of how we observe them (e.g., fixed on the ground, on a balloon, on an aircraft, etc.). In fact, the heading of this section could have been 'From Lagrangian[1] to Eulerian[2] frame of reference' instead.

Let us begin the discussion with a constant background wind that is independent of space and time, i.e., a spatially uniform steady wind. A modification needs to be made on how we describe the phase angle as seen on the ground, thus Eq. 2.3 becomes

$$\begin{aligned} \phi(\vec{r},t) &= \vec{k}\cdot(\vec{r}-\vec{U}^w t)-\Omega t \\ &= \vec{k}\cdot\vec{r}-(\vec{k}\cdot\vec{U}^w+\Omega)t \end{aligned} \tag{2.14}$$

where $\vec{U}^w$ is the wind velocity. Similar to the wave vector $\vec{k}$, $\vec{U}^w$ can be decomposed into three components, x, y and z in the Cartesian coordinates

$$\vec{U}^w = (U^w_x, U^w_y, U^w_z). \tag{2.15}$$

From Eq. 2.14, we can replace $\Omega + \vec{U}^w \cdot \vec{k}$ with ω, introducing the concept of the *ground-based frequency*, which is the frequency observed in the reference frame relative to the ground. The *intrinsic frequency* Ω is the frequency observed within the medium. The terminologies and symbols of these two frequencies are not always consistent between literature and should not be confused. A summary of terminologies used by selected authors can be found in Paper 2, Table 1. Note that the intrinsic frequency of a wave shall not be interpreted as a constant, unchangeable property, even though

[1]Lagrangian description is a description of the fluid motion by following an individual parcel, e.g., a wave packet in our case, as it travels through space and time.

[2]Eulerian description is a description of the fluid motion that fixes with locations in space through time. As in our case, we fixed to the ground and observe the wave.

the word intrinsic might imply this, or confused with the source frequency (an example argument would be 'the intrinsic frequency is constant as long as the wave is generated by the same source'), which will be discussed further in the next section. The intrinsic frequency shall only be understood as the frequency we would observe as if we were moving with the mean flow (e.g., a balloon measurement observes Ω).

Ω and ω are related by the Doppler relationship (Lighthill 1967)

$$\Omega = \omega - \vec{U}^w \cdot \vec{k}, \tag{2.16}$$

where $\vec{U}^w \cdot \vec{k}$ is referred to the Doppler shift term (Buhler 2009). Since it is a dot product between the wind vector and the wave vector, the actual wind component that affects the wave properties seen from the ground is the wind vector projected on the wave vector,

$$U_e^w = \vec{U}^w \cdot \vec{k}/|\vec{k}|, \tag{2.17}$$

where U_e^w is termed the effective wind speed in this thesis.

In the upper atmosphere, the vertical component of the wind velocity can be safely assumed to be negligible compared to the horizontal component, i.e., the horizontal wind component is dominant. Thus, in the rest of this thesis, we will simplify the equations by retaining only the horizontal components of the background wind to study the mean flow interaction with waves. With this assumption, the Doppler relationship is simplified to

$$\begin{aligned} \Omega &= \omega - U_x^w k_x - U_y^w k_y \\ &= \omega - U_e^w k_h, \end{aligned} \tag{2.18}$$

where U_e^w here is the wind component projected on the direction of the horizontal component of the wave vector, k_h.

It is worth emphasising that in the Doppler relationship, the ground-based frequency ω, sometimes called the observed frequency, should be more precisely defined as the frequency observed relative to **the same reference frame in which $\vec{U}^w$ is measured**. This is crucial to keep in mind as we may later encounter wave observations in a different reference frame than the ground (e.g., a moving spacecraft) and we shall properly account for the Doppler shift effect. The analysis becomes even more complicated when we also characterise the wave field in the reference frame fixed to a moving source (e.g., a convection system), which will be discussed further in the subsequent section and in Paper 2.

So far we have discussed the idealistic condition of constant background wind. In the real world, the background wind is neither spatially uniform nor steady. To study the effect of the mean wind, one of the options, though more computationally expensive, is to use a first principle computational fluid dynamics model to simulate the waves in a complicated background situation. Another less expensive option is to apply the ray trace equations to analyse the wavenumbers and the ground-based frequency given the initial conditions (Marks and Eckermann 1995; Heale and Snively 2015). In accordance with Heale and Snively (2015), the ray trace equations read

$$\begin{aligned}
\frac{d\vec{r}}{dt} &= \vec{C}_g \\
\frac{dk_h}{dt} &= -k_h \frac{\partial U_e^w}{\partial x} \\
\frac{dk_z}{dt} &= -k_h \frac{\partial U_e^w}{\partial z} \\
\frac{d\omega}{dt} &= k_h \frac{\partial U_e^w}{\partial t}
\end{aligned} \tag{2.19}$$

where $\vec{C}_g$ is the group velocity which indicates the velocity of a wave packet. An immediate interpretation of the ray trace equations above is that the characteristics of a gravity wave depend on the spatial gradients or rate of change in the effective background wind speed. For instance, a purely vertically varying background flow field modifies the k_z of a wave as a function of z and t, but the horizontal wavenumber and the ground-based frequency keep constant, and thus Ω should be modified due to the dispersion relation.

One of the most important effects of the spatially in-homogeneous mean flow is the channelling of the wave propagation path. Let us still focus on the case of vertically varying horizontal mean flow (or vertical shear flow), If the magnitude of the mean flow is equal to the ground-relative phase speed, i.e., $U_e^w = \frac{\omega}{k_h}$, the intrinsic phase speed $\frac{\Omega}{k_h}$ becomes zero due to the Doppler-shift effect (see Eq. 2.18) and the vertical wavelength λ_z approaches zero (i.e., $k_z \to \infty$) due to the wave dispersion (see Eq. 2.10). When this happens, the gravity wave approaches the *critical level* where the wave can not propagate further up anymore. In contrast, when the intrinsic frequency is Doppler shifted by the mean wind to become close to the local buoyancy frequency of the medium (i.e., $\Omega \to N$), the vertical wavelength λ_z approaches infinity (i.e., $k_z \to 0$) due to the wave dispersion (see Eq. 2.11). This layer is called the *reflection level* or the *evanescent level*, because the wave will be reflected and maybe partially transmitted through this layer. The analysis is similar in the case of a horizontal shear background and we can refer to the modelling study done by Heale and Snively (2015).

The phenomenon of critical level can be clearly examined by a tank experiment shown in Fig. 2.1, where an internal gravity wave was generated by a corrugated wall moving under a tank of stratified fluid. The shadows visualise where the phase lines are within the fluid. The background flow field has a vertical shear which can be verified by the velocity profile superimposed on the photograph (note: measured in the coordinate system moving with the corrugated wall). We can see that the wave train can not penetrate through the critical level as the vertical wavelength becomes infinitely small. Since the vertical wavelength approaches to zero $\lambda_z \to 0$, the vertical shear in the perturbation of horizontal flow speed will become so large that the Richardson number[3] becomes small indicating turbulence production. The increasingly rapid oscillations near the critical region results in *dynamical instability*

[3]Richardson number is the dimensionless number that expresses the ratio of the buoyancy to flow shear.

that leads to wave breaking before reaching the theoretical critical level. Another similar tank experiment is shown in Fig. 2.2 in which the internal waves are initiated by an oscillating cylinder. The waves are generated on both sides of the cylinder, with one propagates to the left and the other to the right in a 2 dimensional flow field. Again, a critical level occurred on the left side for the same reason as in Fig. 2.1. For the wave on the right side, a reflection level occurred due to its intrinsic frequency matched up with the background buoyancy frequency. The reflected wave train interferes with the original wave train, which can be studied in the shadowgraph.

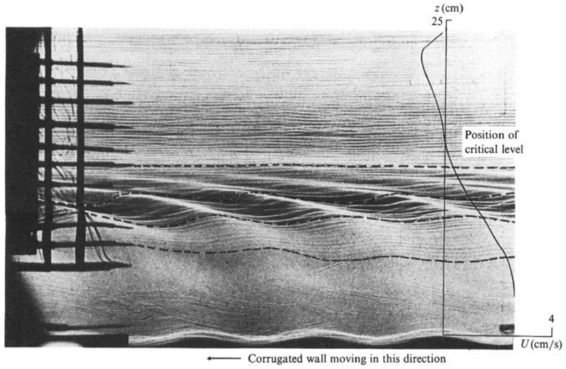

Figure 2.1: A shadowgraph taken from experiments in which an internal wave is generated by leftward-moving set of sinusoidal hills in a stratified shear flow, superimposed a velocity profile of the shear flow measured relative to the velocity of the corrugated wall. [Reproduced, by permission of Cambridge University Press, from Figure 5 in Koop and McGee (1986)]

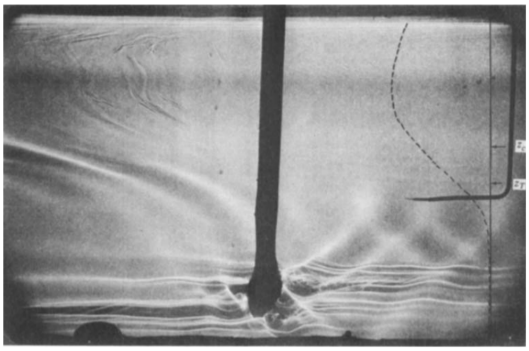

Figure 2.2: A shadowgraph taken from experiments in which internal waves generated by an oscillating cylinder in a leftward-shear flow, superimposed a velocity profile of the shear flow. The leftward-propagating wave encounters a critical level; the rightward-propagating wave encounters a reflection level. [Reproduced, by permission of Cambridge University Press, from Figure 10 in Koop (1981)]

The implications of the critical level and reflection level are significant in the context of atmospheric dynamics. The in-homogeneity of the zonal- and meridional components of the background wind may lead to wave ducting, wave filtering and many other interesting mechanisms. For example, when gravity waves are trapped in between two reflection levels, these waves are prohibited to travel vertically and

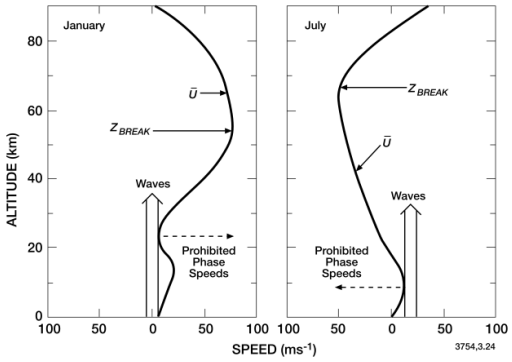

Figure 2.3: A schematic illustration of the wave filtering effect by vertical shear wind in the atmosphere. Left: approximated vertical profile of the mean zonal winds in winter. Right: same as left but in summer. The prohibited and permitted phase speeds of the upward propagating gravity waves and their breaking levels are shown. [Reproduced, by permission of Springer Nature, from Figure 3.11 in G.P Brasseur and Solomon (2005)]

may propagate at large horizontal distance before they break. This wave tunnelling mechanism may act like an 'information link' between non-contiguous geographic regions. Similarly, critical level prohibits gravity waves from propagating upwards at a certain altitude. In fact, the stratospheric zonal winds act like filters to absorb certain waves with the equivalent phase speeds, as illustrated in Fig. 2.3. A seasonal variation also incorporates the pattern of filtering since the stratospheric zonal wind changes its direction with seasons. The remaining waves that propagate higher generate 'wave drag' in the mesosphere, and that in turn produces a mesospheric mean flow pattern, which can be seen in various observations (e.g., Lindzen 1981; Holton 1982). The theoretical studies of wave filtering explain well the observed mesospheric wind pattern that could not be answered previously by pure thermodynamic studies. Besides the seasonal variation in the global wind behaviour, gravity wave filtering also play a key role in the quasi-biennial oscillation (known as 'QBO') and semi-annual oscillation (known as 'SAO') which can be observed in the distributions of chemical compounds.

To give a brief summary, the propagation path of the atmospheric internal wave depends on the spatial distributions of the background wind due to the Doppler effect and wave dispersion. In turn, the results of wave ducting and filtering influence where the wave dissipation occurs in the atmosphere, especially in the mesosphere. However, the origin and magnitude of the wave dissipation still remains an active topic for debate in current studies (Fritts and Alexander 2003; Vadas and Becker 2018; Vadas, Zhao, et al. 2018; B. Kaifler et al. 2015; N. Kaifler et al. 2017).

2.2.2 Gravity wave sources

In this section, we will investigate the Doppler relationship (Eq. 2.16) one step further to incorporate the consideration of the frame of reference that is moving with a gravity wave source. In fact, an internal gravity wave is generated by an obstacle having a relative motion with respect to the ambient medium forcing the

fluid parcels to oscillate. As expected from the towing tank experiment shown earlier in Fig. 2.1 (Koop and McGee 1986), the intrinsic frequency, Ω, can also be expressed in terms of relative velocity, $\vec{U}^{rel}$, between the source and the background wind

$$\begin{aligned} \Omega &= -\vec{U}^{rel} \cdot \vec{k} \\ &= -(\vec{U}^{w} - \vec{U}^{s}) \cdot \vec{k}, \end{aligned} \tag{2.20}$$

where $\vec{U}^{s}$ is the source velocity with respect to the same frame of reference as $\vec{U}^{w}$, e.g., relative to the ground. Here $\vec{U}^{rel}$ can also be seen as the wind velocity relative to reference from fixed to the (moving) source. By relating Eq. 2.16 and Eq. 2.20, one can conclude that the ground-based frequency is

$$\omega = \vec{U}^{s} \cdot \vec{k}. \tag{2.21}$$

Let us examine the implication of Eq. 2.20 and Eq. 2.21 with an hypothetical experiment. The first scenario is an uniform flow with a certain speed over a stationary obstacle (i.e., an obstacle not moving relative to the ground). The second scenario is a moving corrugated wall with the same speed but opposite direction in a stationary medium, in analogy to the laboratory experiment made by Koop and McGee (1986). Assuming that the ambient stratification and the shapes of the obstacles are identical in both scenarios, internal waves with the same spatial characteristics (e.g., k_x, k_z, etc.) will be excited, hence with the same intrinsic frequency Ω (recall the dispersion relation given in Eq. 2.11). The differences between these two scenarios are that, in scenario 1 a stationary observer, hence fixed to the ground and the source in this case, will record a stationary wave with $\omega = 0$ and $\Omega = -\vec{U}^{w} \cdot \vec{k}$; while in scenario 2 a transient wave will be observed by a stationary observer with $\omega = \Omega = \vec{U}^{s} \cdot \vec{k}$. A graphical illustration of this hypothetical experiment is given in Paper 2 Fig. 2.

In Eq. 2.21, the connection between the ground-based frequency and the background wind vanishes. This is because the ground-based frequency ω will not be Doppler-shifted by the velocity of the medium under steady flow conditions (i.e., not change in time), but by the speed of the obstacle with respect to the observer. This is similar to the classical example of the propagating sound wave Doppler-shifted by a moving ambulance.

The kind of towing experiment generates 'lee waves', internal gravity waves that are phase-locked with the motion of a disturbance. Hence, the phase speed observed by a ground-based instrument is equal to the speed of the obstacle that excites the wave in this way, which was termed the 'obstacle effect' (Clark et al. 1986). This excitation mechanism is clearly examined by Prusa et al. (1996) who performed a numerical study to approximate a travelling tropospheric forcing. Later Fovell et al. (1992) made a more comprehensive numerical simulation to study the gravity wave excitation by a convective system, where they found that, in addition to the 'obstacle effect', there is another mechanism primary excites gravity waves in the absence of storm-relative mean winds, that was labelled as the 'mechanical oscillator effect' by Clark et al. (1986). This mechanism is closely analogous to the laboratory tank experiment, such as the one shown in Fig. 2.2 (Koop 1981), where a constant

oscillating cylinder generates waves in each of the quadrants in a 2-dimension flow field with phase lines forming a fixed angle of β that satisfies $\cos\beta = \Omega/N$ (Eq. 2.13).

More recent studies focus on the so-called 'secondary gravity waves' excited by local body forces in fluid medium, such as the studies made by Vadas, Zhao, et al. (2018) and Vadas and Becker (2018), where they examine the theoretical basis of this type of gravity waves supported by numerical simulations and observational evidences in the atmosphere. The responses of spontaneous local forces is an initiation of a full spectrum of gravity waves, which shall be differentiated from the previously mentioned mechanisms, the 'obstacle effect' and 'oscillator effect', where a monochromatic (set of) wave(s) is (are) generated in an idealistic background condition. In a laboratory environment, this type of internal waves may be excited by a single 'tap' on a cylindrical oscillator in a tank of stratified fluid instead of giving a constant oscillating frequency. In such condition, a wave field with phase angles radiate in all directions can be observed.

In the context of atmospheric gravity waves, a summary of commonly investigated sources are, among others, tomographic generation (e.g., mountains), convective generation (e.g., convection systems), shear generation (e.g., Kelvin-Helmholtz instability), geostrophic adjustment (e.g., baroclinic instability), wave-wave interactions, etc. The detailed classifications and an extensive list of references to the original literature can be found in Fritts and Alexander (2003).

In summary, the characterisation of gravity waves depends on both the state of the ambient medium (e.g., stratification, mean flow) and excitation mechanisms by the source. The observable wave parameters are govern by theoretical formulations based on linear wave theory. The foundation of theory must be applied correctly when we attempt to interpret the data obtained by observations. In the following section we will look at how these theoretical formulations can be used to further derive unknown wave parameters for different types of measurement technique.

2.3 Observations and interpretation

Internal gravity waves introduce propagating disturbances in the medium, not only changing the local velocity field, $U_x(\vec{r})$, $U_y(\vec{r})$, $U_z(\vec{r})$, but also the atmospheric temperature field, $T(\vec{r})$ and density field, $\rho(\vec{r})$. Consequently, instruments that can detect these physical observables in the atmosphere can be used for gravity wave analysis by subtracting their mean component from the total field as

$$X' = X - \bar{X}, \tag{2.22}$$

where X denotes an arbitrary observed quantity that can be decomposed into a mean component, often taken to be an average over time and/or space, $\bar{X}$ and a perturbation component X'. The exact method used to estimate the mean value may vary, and several different filtering techniques are used by different studies (e.g., Ehard et al. 2015). Although it is important to keep in mind that the choice of filter may influence the resulting spectrum of the gravity wave data that is obtained, the

discussion around the filtering technique is out of the scope of this thesis. Instead, we will focus on how different observational geometry influences what wave parameters that can be retrieved directly or indirectly.

Recent developments in remote sensing techniques allows us to capture small-scale features even in the middle and upper atmosphere. Each technique has its own advantage for characterising gravity waves and they are often complementary to one another. Airborne and space-borne instruments are commonly used to take spatial snapshots of structures at high altitudes with great potential for mapping the global distribution of gravity wave (Wu et al. 2006; Preusse, Eckermann, et al. 2008) while ground-based instruments are often used to take time series, with the advantage of high temporal resolution at a single site (Gardner and Taylor 1998; Franzen et al. 2018).

As mentioned earlier in the theoretical section (Sect. 2.2), gravity waves are characterised by their fundamental properties: wavelength, frequency and amplitude of the fluctuations in atmospheric state. These observable characteristics of the waves are governed by the dispersion relation and the Doppler relation, thus they are strongly dependent on spatial and temporal variations of the mean wind, as well as the properties of the sources that generate the waves. All wave parameters (vertical-, horizontal wavelength, frequency and amplitude) can rarely be measured directly using a single instrument. Additional information, among others, such as the time averaged background wind is needed to retrieve the desired wave properties indirectly.

In this section, a consistent framework is described for deriving monochromatic wave parameters with the consideration of the background winds, as well as the generation source speed under the assumption that the wave is generated by the 'obstacle effect' mechanism. This source speed is essentially the phase speed of the wave. In order to simplify the analysis, the 'mechanical oscillator effect' is not in consideration here in this thesis, as its analysis involves yet another Doppler relation between the source frequency and the observed frequency (note that the aforementioned Doppler relation generally refer to the relationship between the intrinsic frequency and the ground-based frequency). Also, without affecting the purpose of illustration, such as the ones shown in Fig. 2.4 and Fig. 2.5, the dispersion relation used in this section is Eq. 2.11 and the Doppler relation is Eq. 2.18.

The framework can be applied to three different types of observations: time series imagers, 3D imagers and ground-based lidars. By following the framework, observations of gravity waves by different techniques can be compared more directly. The framework also provides an overview of what additional knowledge is needed to analyse the desired quantities when treating quasi-monochromatic waves. Note that we do not consider the systematic limitations of particular instruments in this section, these are described in detail by, e.g., (Gardner and Taylor 1998).

2.3.1 Imager observations

Imager measurements (spatial snapshots) can be roughly categorised into two types, (1) vertical-horizontal imagers such as limb-measurements from space and (2)

horizontal-horizontal imagers such as zenith sky measurements from the ground or nadir-measurements from space.

The vertical-horizontal type of imagers can be used to obtain the vertical wavelength. However, the horizontal wavelength can only be assumed to a certain extent because the imaging plane is not always perpendicular to the wave front in three dimensional space (Preusse, Dörnbrack, et al. 2002). An improvement can be made by using 3D reconstruction to obtain the 'true' horizontal wavelength (Ungermann 2011). Several new imaging instruments use the 3D tomographic method such as (Larsson et al. 2015; Song et al. 2017; Krisch et al. 2017). The intrinsic- and ground-based period can be calculated from the dispersion relation and the Doppler relation, respectively, with the help of the additional knowledge about the background wind. Once the ground-based frequency and the horizontal wavelength are known, we can estimate how rapidly the source is moving. The framework for 3D imagers is presented in the left panel of Fig. 2.4.

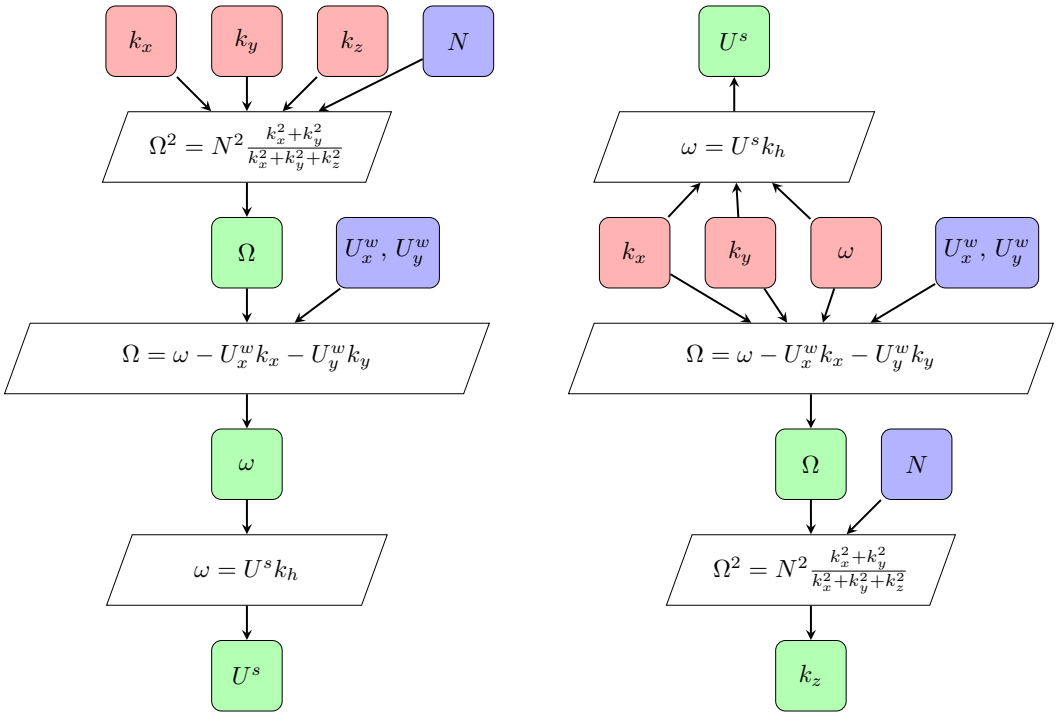

Figure 2.4: Frameworks for 3D imagers (left) and time series imagers (right). Observed wave properties are coloured in red, calculated properties are in green and additional information, i.e., background wind in blue.

The horizontal-horizontal type of imagers are able to measure the horizontal wavelength and its orientation (e.g., Armstrong 1982; Nakamura et al. 1999; Medeiros et al. 2003). Ground-based phase velocity can be estimated by recording a time series at a fixed position, so the ground-based frequency is obtained. It is a challenge to be sensitive to the vertical wavelength because of the measurement geometry. Nevertheless, again, with additional information on the background wind, the vertical

wavelength and intrinsic frequency can be calculated in a similar manner by applying the Doppler relation and the dispersion relation. In addition, the source speed can be obtained, rather straight forwardly, from the measured horizontal wavelength and ground-based frequency, as illustrated in the right panel of Fig. 2.4.

2.3.2 Ground-based vertical profilers

Ground-based lidar observations provide high temporal and vertical resolutions (Baumgarten et al. 2015). It can be used to determine the vertical wavelength and the ground-based frequency. Especially a stationary wave ($U^s = 0$) can be identified from the vertical-time series if the phase lines are nearly horizontal ($\omega = 0$). Unfortunately, since ground-based lidar does not provide horizontal spatial information, k_h can not be easily estimated.

Firstly, additional knowledge about the orientation of the wave propagation is needed to properly account for the effective wind speed, U_e^w, so that the Doppler relationship (Eq. 2.18) can be used. Knowing this, the horizontal wavelength is still ambiguous when one attempts to solve the quartic equation by combining the dispersion relation and the Doppler relation as illustrated in the right panel of Fig. 2.5.

Secondly, if we know the speed of the moving source, in theory, it is possibly to derive the horizontal wavelength by using Eq. 2.21. U^s can be estimated, for instance, by weather radar if a wave is generated by convective systems, or the phase speed of the primary wave when a secondary wave has been generated. Once the horizontal wavelength is determined, the intrinsic frequency can be calculated from the dispersion relation and the effective wind speed can be obtained from the Doppler relationship (Eq. 2.18). The proposed frameworks are illustrated in the left panel of Fig. 2.5.

To conclude, in all the observation types discussed above, knowledge about the background winds or the phase speed of the wave is crucial if one is to attempt to determine additional wave parameters by applying the linear wave theory described in Sect. 2.2. Furthermore, regardless of the direction of the local background wind, the orientation of the quasi-monochromatic wave can be in all directions. Since the Doppler relation Eq. 2.18 involves dot product of the wave vector and the wind vector we need to know their relative directions. In other words, effective background wind speed, U_e^w in Eq. 2.17 needs to be obtained.

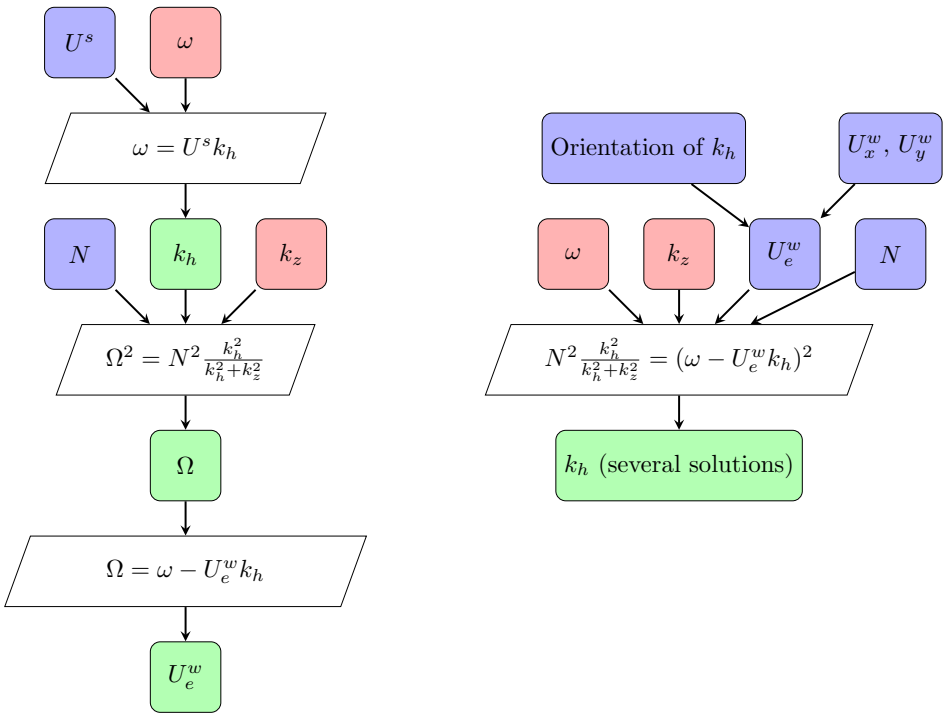

Figure 2.5: Frameworks for ground-based lidar observations. Observed wave properties are coloured in red, calculated properties are in green and additional information in blue. The left panel shows when U^s is known. The right panel shows when U_e^w is known and the resulting horizontal wavelength may be up to 4 real solutions due to the quartic function.

Chapter 3

Mesospheric photochemistry

3.1 Introduction

Almost all of the atoms and molecules in the mesosphere participate in chemical processes, which result in their spatial and temporal distributions being closely linked to one another. Processes that involve the interaction with light are called photochemical processes. In the field of aeronomy, the source of light we are typically concerned with is the sun. The ultra-violet photons in the solar spectrum have sufficient energy to dissociate molecules (i.e., photo-dissociation), or may excite molecules from a basic ground state to an energetic state (i.e., photo-excitation). These excited molecules are often unstable and eventually spontaneously release their "excess" energy as photons (i.e., fluorescence), or by colliding with other molecules to transfer this energy (e.g., inter-molecular energy transfer, quenching). Ozone, atomic- and molecular oxygen are typical examples of the constituents in the mesosphere that are constantly dealing with these types of energy transfer originating in the sun. In fact, thanks to these photochemical processes that act as a protection layer for us, the most powerful solar energy is mostly attenuated before reaching the lower atmosphere.

Among the above mentioned photochemical processes, the photon energy that is spontaneously emitted by the electronically excited chemical species in the atmosphere is called airglow. For instance, Fig. 3.1 shows a thin greenish layer on the night side of Earth's atmosphere that can be observed by human eyes from space. Depending on if the solar energy is available or not when it is observed, airglow is called either dayglow or nightglow. Common airglow emissions are from excited OH , NO, Na, Li and O_2. This thesis addresses airglow emissions from O_2 at wavelengths around 762 nm and 1.27 µm in the mesospheric region.

One of the important motivations for measuring airglows is that they paint the atmospheric gravity waves that are present in the atmosphere. Since the kinetics of the airglow chemistry (described in Sect. 3.3) involves numerous rate coefficients that are sensitive to temperature, as well as to the density of photochemically stable species (e.g., O_2, N_2) that collide with the unstable species, any fluctuation in the atmospheric temperature and background density will affect the density distributions of the excited species. In turn, the observed airglow emissions often exhibits a 'wavy'

Figure 3.1: Airglow above the horizon captured from the International Space Station. Image from NASA.

Figure 3.2: Airglow exhibits in a stripped pattern observed from the ground over Maine, USA. [Permitted by Mike Taylor and Sonia MacNeil]

pattern such as the observation shown in Fig. 3.2. Therefore, monitoring the airglow layer can be used as a great tool to study not only the morphology but also the distribution of gravity waves (see Chapter 2) exist in the mesopause region.

Moreover, the observation of the oxygen airglow is also motivated by the close connection to the concentration of ozone molecules that are present in the mesosphere. Typically, the ozone concentration is measured by observing their thermal emission, such as in the spectral region around 9.6 µm and in microwave region, or by measuring their absorption in the UV-visible region, such as limb-scatter technique and solar/stellar occultation techniques. However, the density of ozone in the mesosphere changes drastically from the nighttime to the daytime, with as little as only 10 % during daytime of the maximum values observed in the nighttime. In addition, the total density in the mesosphere is much lower than in the stratosphere because of the exponential decay of the density distribution with altitude. This means that measurements of the daytime ozone density in the mesosphere require instruments with high sensitivity. Fortunately, as previously mentioned, oxygen airglows serve as useful tool to infer the ozone density.

In this chapter, we will start by listing the necessary information on chemical kinetics concepts as well as some of the important nomenclature that is used to implement a photochemical model. After that, we will discuss a general kinetic

model that describes the oxygen airglow photochemistry. This photochemical model is the foundation of uncovering the important linkage between the ozone photolysis and the density of excited oxygen molecules.

3.2 Fundamental chemical kinetics

The concern of chemical kinetics is to ask questions on how fast a chemical reaction is supposed to occur. Since almost all of the minor constituents in the atmosphere are created by chemical processes, the concentration of each chemical species highly depends on the rates at which they are produced and removed. These reaction rates as parameters describing the kinetics of the reactions must be determined in the laboratory, typically by recording the concentration of the reactants or products as a function of time at a controlled background conditions. A discussion about the measurement techniques and the associated uncertainties of these measured parameters is out of the scope of this thesis. This section is aimed to provide necessary information at an elementary level for us to be able to understand some of the important nomenclature, as well as how the implementation of a photochemical model is made. Such a model will be presented in the Sect. 3.3.

3.2.1 The steady state assumption

Let us begin with a general expression of a reaction between species A, B, C,...,

$$a\,\mathrm{A} + b\,\mathrm{B} + c\,\mathrm{C} + ... \xrightarrow{k} e\,\mathrm{E} + f\,\mathrm{F} + g\,\mathrm{G} + ... \tag{3.1}$$

where lower case letter represent integers that indicate the stoichiometry and k is the *rate coefficient* or *rate constant* for this particular reaction, which often has a dependency on the local temperature.

The rate of change of a particular reactant or of a particular product is expressed as

$$R = -\frac{1}{a}\frac{\mathrm{d[A]}}{\mathrm{d}t} = -\frac{1}{b}\frac{\mathrm{d[B]}}{\mathrm{d}t} = ... = \frac{1}{e}\frac{\mathrm{d[E]}}{\mathrm{d}t} = \frac{1}{f}\frac{\mathrm{d[F]}}{\mathrm{d}t} = ... \tag{3.2}$$

The reaction rate depends on the concentrations of reactants so the rate equation often, but not always, takes the form of

$$R = k[\mathrm{A}]^a[\mathrm{B}]^b[\mathrm{C}]^c... \tag{3.3}$$

where square brackets denote the concentration of species, which usually has a unit of molec cm^{-3}, so k has molec $\mathrm{cm}^{-3\mathrm{x}}\mathrm{s}^{-1}$ units according to the order of the reaction.

In an environment like the atmosphere, where many species are reacting with each other simultaneously, chemical reactions often occur through an interconnected network called *reaction mechanism*. In such case, the rate of change of a certain species must consider all the possible production and loss processes. For example, a

reaction mechanism represented by

$$\mathrm{A} + \mathrm{B} \xrightarrow{k_1} \text{Products} \tag{3.4}$$

$$\mathrm{A} + \mathrm{C} + \mathrm{M} \xrightarrow{k_2} \text{Products} \tag{3.5}$$

$$\mathrm{A} + \mathrm{F} \xrightarrow{k_3} \text{Products} \tag{3.6}$$

$$\mathrm{G} + \mathrm{H} \xrightarrow{k_4} \mathrm{A} + \text{Products}. \tag{3.7}$$

If we focus on the species A, the sum of all the production terms (also called sources) P_i is

$$\sum_i P_i = k_4[\mathrm{G}][\mathrm{H}], \tag{3.8}$$

and the sum of all the losses (also called sinks) of species A $L_i[\mathrm{A}]$ is

$$\sum_i L_i[\mathrm{A}] = (k_1[\mathrm{B}] + k_2[\mathrm{C}][\mathrm{M}] + k_3[\mathrm{F}])[\mathrm{A}]. \tag{3.9}$$

The rate of change of species A results in

$$\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = \sum_i P_i - \sum_i L_i[\mathrm{A}] \tag{3.10}$$

When the density of A is not changing over time, i.e., $\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = 0$, the constituent is said to be in the *steady state*. This means that whenever A is newly produced it will be instantaneously removed by other reactions of the network so that A will not be accumulated in the system. Thus the concentration of A under the steady state assumption can be derived as

$$[\mathrm{A}] = \frac{\sum_i P_i}{\sum_i L_i} = \frac{k_4[\mathrm{G}][\mathrm{H}]}{k_1[\mathrm{B}] + k_2[\mathrm{C}][\mathrm{M}] + k_3[\mathrm{F}]} \tag{3.11}$$

In a chemical model which includes multiple reaction mechanisms, such as the one described in Sect. 3.3, the steady state assumption is particularly useful to derive the concentrations of all species considered in the model. However, if the time scale of the dynamics in a certain location is comparable to the lifetime of a certain species, i.e., $\tau = 1/\sum_i L_i$, the effect of transport must be taken into account in the balance equation (i.e., the continuity equation)

$$\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = \sum_i P_i - \sum_i L_i[\mathrm{A}] - \nabla \cdot ([\mathrm{A}]\vec{v}) \tag{3.12}$$

where the last term added represents the advection effect on the species A by the mass flow which has a velocity $\vec{v}$. We will not discuss further the details about the effect of transport in this thesis, but shall be aware of the bias introduced by the assumption of only the chemical steady state.

3.2.2 Photochemical processes

A photochemical reaction is defined as a chemical reaction that concerns the interaction with electromagnetic energy. The reactants and products of the processes often associate with excited states (which was previously referred to having "excess energy" in the introduction section of the chapter). Depending on the amount of energy that is absorbed, the excited state can be electronically, vibrationally and rotationally excited. In this thesis, we will mainly consider electronically and, to a lesser extend, vibrationally excited states.

As briefly mentioned earlier in the introduction section, photo-excitation can be the first step of a chain of photochemical processes. This is also termed *resonance absorption*, or in the context of atmosphere, *solar-excitation* since the sun is considered as the only energy source. As an example, the reaction of the solar-excitation of molecule XY takes the form of

$$\mathrm{XY} + \mathrm{h}\nu \xrightarrow{\mathrm{g}} \mathrm{XY}^* \tag{3.13}$$

where the superscript * denotes that the molecule is in an excited state, g is the excitation rate which has a unit of s^-1. If the associated photon energy $\mathrm{h}\nu$ (Planck's constant times the photon frequency) is sufficiently high, the molecule XY may fall apart to form multiple constituents which is referred to as *photo-dissociation*, or *photolysis* process, represented by,

$$\mathrm{XY} + \mathrm{h}\nu \xrightarrow{\mathrm{J}} \mathrm{X} + \mathrm{Y}^* \tag{3.14}$$

note that either, both or neither of the products can be in an excited state. Here, J is termed photolysis rate. Since the reaction only involves one reactant, the photolysis rate has a unit of s^{-1}.

Both excitation rate g and photolysis rate J are determined by how many photons are available to interact with the gas, the ability of the particular species to absorb these photons and the probability that the photon leads to the specific products of the photochemical reaction. This means that to compute these rates, we need radiative transfer modelling along the path from the light source to the location of interest for each wavelength. In other words, the solar irradiance, the absorption cross-sections of the species and the quantum yield for the particular photo-dissociation are needed to proceed with the computation of g and J.

After the initiation of an excited molecule, the molecule may transfer its energy to another molecule by collision, which is called the *inter-molecular energy transfer* process, i.e.,

$$\mathrm{XY}^* + \mathrm{AB} \xrightarrow{\mathrm{k}} \mathrm{XY} + \mathrm{AB}^* \tag{3.15}$$

or may collide with another molecule and simply returns back to its basic ground state, which is referred to the *quenching* process, i.e.,

$$\mathrm{XY}^* + \mathrm{M} \xrightarrow{\mathrm{k}} \mathrm{XY} + \mathrm{M}. \tag{3.16}$$

Another way of returning back to the basic ground state without collision is by *fluorescence* or *chemiluminescence* process (naming depends on how the molecule

was excited), that is spontaneously releasing the photon energy at a specific spectral characteristic, i.e.,

$$\mathrm{XY}^* \xrightarrow{\mathrm{A}} \mathrm{XY} + \mathrm{h}\nu, \tag{3.17}$$

where A is termed the Einstein A coefficient, as known as the inverse of photochemical lifetime, i.e., $A = 1/\tau_{\mathrm{XY}^*}$. Similarly to the reaction coefficients k, the Einstein A coefficient must be obtained by laboratory measurements. This spontaneous relaxation of the excited state is exactly the reason why we can observe the airglow by measuring the released photo energy $\mathrm{h}\nu$ in a specific spectral region. This emitted photon energy may be re-absorbed to repeat the excitation process, which is commonly referred to as *self-absorption*.

In summary, we have mentioned some fundamental chemical kinetic concepts that are important for implementing a photochemical model, as well as introduced most of the nomenclature that will be used in Sect. 3.3 where we discuss the photochemical model that describes the relationship between ozone and several excited state of oxygen in the mesosphere.

3.3 Oxygen airglow photochemistry

As mentioned in Chapter 1, the current Odin satellite includes an instrument OSIRIS onboard. The optical spectrograph (hereafter OS) routinely measures the oxygen atmospheric band (A-band) spectra centred at 762 nm, and the infrared imager (hereafter IRI) measures the oxygen infrared atmospheric band (IRA-band) intensity centred at 1.27 µm. In addition, the future satellite MATS will also capture the A-band emission. Oxygen A-band and IRA-band emissions are the conventional names for the airglow emissions by the two electronically excited states of the oxygen molecule, $\mathrm{O}_2(\mathrm{b}^1\,\Sigma_g{}^+)$ and $\mathrm{O}_2(\mathrm{a}^1\,\Delta_g)$, respectively. Both of them are closely connected to the available ozone amount in the mesosphere under sunlit conditions, since they are the 'by-products' of the ozone photolysis in the Hartley band (around 310 nm to 350 nm, see Fig. 3.4). For this reason, the observed intensities of these two oxygen band emissions during the day can often be utilised as proxies for the ozone density in the mesosphere (Valentine A. Yankovsky et al. 2016; Thomas et al. 1984; M. G. Mlynczak et al. 2001; Martin G. Mlynczak et al. 2007; P. Sheese 2009). In this section, we will investigate the most important reactions that connect ozone and the several electronically excited states of the oxygen atom and molecule. The addition of their vibrational-rotational sub-levels can be referred to the model in V A Yankovsky and Manuilova (2006) and Valentine A. Yankovsky et al. (2016) and in Paper 1. The kinetic scheme of the photochemical model considered here is illustrated in Fig. 3.3.

The entire reaction chain starts by ozone molecules being photo-dissociated by the solar energy in the Hartley band (see Fig. 3.4 for the spectral range)

$$\mathrm{O}_3 + \mathrm{h}\nu \xrightarrow{J_H} \mathrm{O}(^1\mathrm{D}) + \mathrm{O}_2(\mathrm{a}^1\Delta_g), \tag{3.18}$$

where J_H is the photolysis rate in this spectral range. While the theoretical combination of the products (an oxygen atom and a molecular oxygen) of this particular

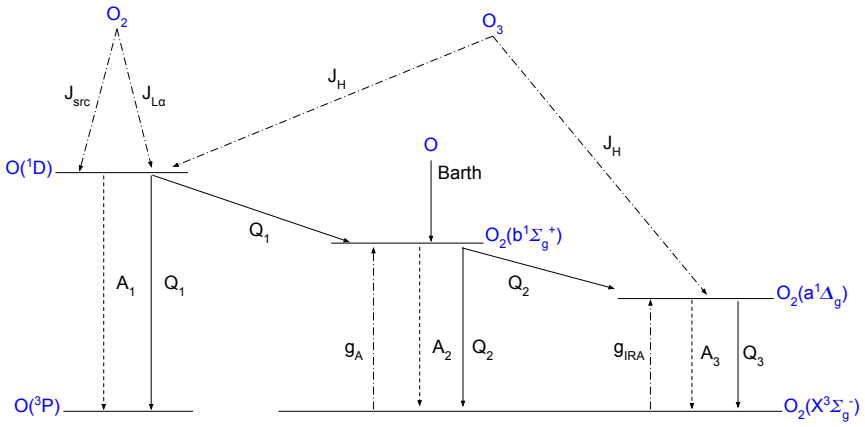

Figure 3.3: Scheme of kinetics that is considered in this section. Dashed-dotted lines represent either photolysis or resonance absorption, dashed lines represent the spontaneous de-excitation processes and solid lines represent quenching processes.

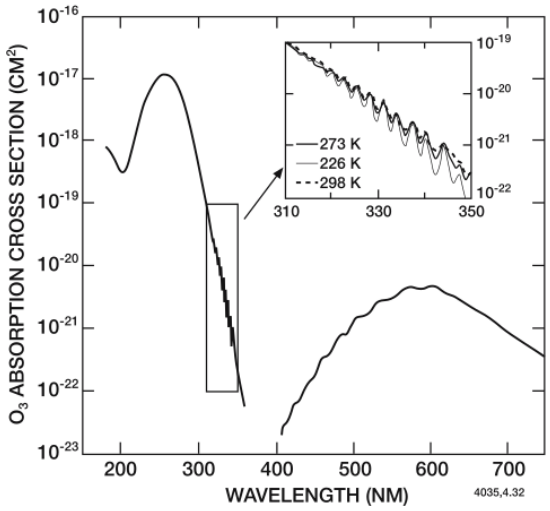

Figure 3.4: The ozone cross section in the Hartley band (200 nm to 300 nm), Huggins bands (310 nm to 350 nm) and Chappuis bands (410 nm to 750 nm). [Reproduced, by permission of Springer Nature, from Figure 4.35 in G.P Brasseur and Solomon (2005)]

photochemical reaction depends on the wavelength of the incident photo energy, $O(^1D)$ and $O_2(a^1\Delta_g)$ are the most probable products and have a quantum yield of 90 % based on laboratory experiment. Besides the Hartley band, other absorption bands of ozone are in longer wavelength region, such as Huggins bands and Chappuis bands, as shown in Fig. 3.4. Although the Huggins bands and Chappuis bands are not in the focus on our airglow photochemistry, it is worth mentioning that they are routinely measured by the Odin mission (more specifically, the optical spectralgraph OSIRIS).

Another important source of the excited oxygen atom $O(^1D)$, besides the photol-

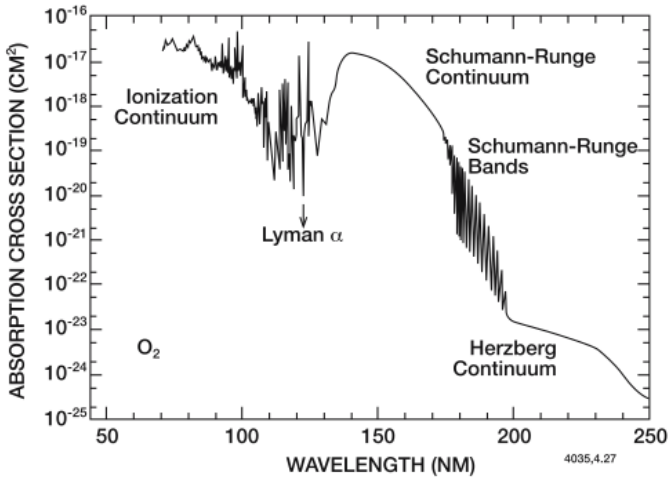

Figure 3.5: The molecular oxygen cross section. [Reproduced, by permission of Springer Nature, from Figure 4.30 in G.P Brasseur and Solomon (2005)]

ysis of ozone, comes from the photo-dissociation of molecular oxygen

$$\mathrm{O_2} + \mathrm{h}\nu \xrightarrow{J_{src},\, J_{L\alpha}} \mathrm{O} + \mathrm{O}(^1\mathrm{D}), \tag{3.19}$$

where J_{src} and $J_{L\alpha}$ are the photolysis rates corresponding to the Schumann-Runge Continuum and Lyman α line, respectively (see Fig.3.5 for spectral range). Here, chemical species without brackets denote their ground states, for instance $\mathrm{O_2}$ is equivalent to $\mathrm{O_2}(\mathrm{X}^3\Sigma_g^-)$ and O is $\mathrm{O}(^3\mathrm{P})$. The absorption cross section of molecular oxygen is shown in Fig. 3.5.

The excited atomic oxygen $\mathrm{O}(^1\mathrm{D})$ then either spontaneously release its energy, or transfer its energy by quenching processes

$$\mathrm{O}(^1\mathrm{D}) \xrightarrow{A_1} \mathrm{O} + \mathrm{h}\nu \tag{3.20}$$

$$\mathrm{O}(^1\mathrm{D}) + \mathrm{N_2}, \mathrm{O_2} \xrightarrow{k_1^{N_2},\, k_1^{O_2}} \text{all products} \tag{3.21}$$

where subscript $_1$ of the reaction rate coefficients corresponds to the quenching series represented in Fig. 3.3, likewise for the Einstein A coefficient. The superscripts ^{O_2} and ^{N_2} denote the rate coefficients corresponding the quenching by the species. Hence, the concentration of $\mathrm{O}(^1\mathrm{D})$ can be derived by assuming the sources and the sinks are balanced (i.e. photochemical steady state), i.e.,

$$[\mathrm{O}(^1\mathrm{D})] = \frac{(\phi_{src} J_{src} + \phi_{L\alpha} J_{L\alpha})[\mathrm{O_2}] + \phi_H J_H [\mathrm{O_3}]}{A_1 + Q_1}, \tag{3.22}$$

where ϕ_{src}, $\phi_{L\alpha}$ and ϕ_H denote the quantum yields of each of the photolysis processes, and

$$Q_1 = k_1^{N_2}[\mathrm{N_2}] + k_1^{O2}[\mathrm{O_2}]. \tag{3.23}$$

Among the two quenching processes represented in reaction 3.21, the most important reaction in our kinetic scheme is the collision with O_2 that partially forms the excited oxygen molecule $O_2(b^1\Sigma_g{}^+)$, i.e.,

$$O(^1D) + O_2 \xrightarrow{\beta k_1^{O_2}} O + O_2(b^1\Sigma_g{}^+), \tag{3.24}$$

where β represents the fractional efficiency of this product path. Another important source of $O_2(b^1\Sigma_g{}^+)$ is the solar-excitation at the A-band wavelengths of the oxygen molecule itself, i.e.,

$$O_2 + h\nu \xrightarrow{g_A} O_2(b^1\Sigma_g{}^+), \tag{3.25}$$

where g_A is the A-band excitation rate. A small contribution to the production of $O_2(b^1\Sigma_g{}^+)$ is by the two-step transfer mechanism, as known as the Barth-type mechanism, represented by (after McDade et al. (1986))

$$O + O + M \xrightarrow{\alpha k^O} O_2{}^* + M \tag{3.26}$$

$$O_2{}^* + O_2 \xrightarrow{\gamma k_*^{O_2}} O_2(b^1\,\Sigma_g^+) + O_2, \tag{3.27}$$

$$O_2{}^* + O_2, N_2, O \xrightarrow{k_*^{N_2}, k_*^{O_2}, k_*^O} \text{all products}, \tag{3.28}$$

$$O_2{}^* \xrightarrow{A_*} O_2 + h\nu, \tag{3.29}$$

$$O_2(b^1\Sigma_g^+) + O_2, N_2, O \xrightarrow{k_2^{N_2}, k_2^{O_2}, k_2^O} \text{quenched products}, \tag{3.30}$$

$$O_2(b^1\Sigma_g^+) \xrightarrow{A_2} O_2 + h\nu, \tag{3.31}$$

where α and γ represent the fractional efficiency of the three-body recombination into the unspecified excited state $O_2{}^*$ and from this state to $O_2(b^1\Sigma_g^+)$, respectively. The subscript $_*$ denotes the Einstein A coefficient and the quenching rate coefficients of the unspecified excited state $O_2{}^*$. In photochemical equilibrium, the concentration of $O_2(b^1\Sigma_g^+)$ solely resulted from the Barth-type mechanism can be written as

$$[O_2(b^1\Sigma_g^+)]_{Barth} = \frac{\alpha k^O [O]^2 [M] \gamma k_*^{O_2} [O_2]}{(A_2 + Q_2) \cdot (A_* + Q_*)}, \tag{3.32}$$

where

$$\begin{aligned} Q_2 &= k_2^{O_2}[O_2] + k_2^{N_2}[N_2] + k_2^O[O], \\ Q_* &= k_*^{O_2}[O_2] + k_*^{N_2}[N_2] + k_*^O[O]. \end{aligned} \tag{3.33}$$

However, α, γ, A_* and k_* are quantities that are often unknown and difficult to measure, thus introducing empirical quenching coefficients can simplify the problem as (D. P. Murtagh et al. 1990)

$$[O_2(b^1\Sigma_g^+)]_{Barth} = \frac{k_1 [O]^2 [M] [O_2]}{(Q_2 + A_2) \cdot (C^{O_2}[O_2] + C^O[O])}, \tag{3.34}$$

where

$$C^{O_2} = \frac{1 + Rk_*^{N_2}/k_*^{O_2}}{\alpha\gamma}, \qquad C^O = \frac{k_*^O/k_*^{O_2}}{\alpha\gamma}. \tag{3.35}$$

These newly introduced empirical quenching coefficients C^{O_2} and C^O were evaluated by rocket measurements of nightglow emissions (D. P. Murtagh et al. 1990). Although the Barth-type mechanism is not a significant source compared to other sources to form $O_2(b^1\Sigma_g^+)$, it is special in a way that, it does not involve absorption of solar radiation, which naturally becomes the only source in the nighttime. Nevertheless, summing up all the sources, the concentration of $O_2(b^1\Sigma_g^+)$ can be derived as

$$[O_2(b^1\Sigma_g^+)] = \frac{\beta k_1^{O_2}[O(^1D)][O_2]}{A_2 + Q_2} + \frac{g_A[O_2]}{A_2 + Q_2} + [O_2(b^1\Sigma_g^+)]_{Barth} \tag{3.36}$$

The quenching processes of $O_2(b^1\Sigma_g^+)$ upon collisions (in reaction 3.30) result in the production of $O_2(a^1\Delta_g)$, e.g.,

$$O_2(b^1\Sigma_g^+) + N_2, O_2, O \xrightarrow{\epsilon k_2^{N_2}, \epsilon k_2^{O_2}, \epsilon k_2^O} O_2(a^1\Delta_g) + N_2, O_2, O, \tag{3.37}$$

where ϵ is the fractional efficiency of the quenching processes that actually produce $O_2(a^1\Delta_g)$.

Similar to $O_2(b^1\Sigma_g^+)$, $O_2(a^1\Delta_g)$ can also be produced by resonant absorption from the ground state, removed by spontaneous emission and quenching processes, i.e.,

$$O_2 + h\nu \xrightarrow{g_{IRA}} O_2(a^1\Delta_g), \tag{3.38}$$

$$O_2(a^1\Delta_g) \xrightarrow{A_3} O_2 + h\nu, \tag{3.39}$$

$$O_2(a^1\Delta_g) + N_2, O_2, O \xrightarrow{k_3^{N_2}, k_3^{O_2}, k_3^O} O_2 + N_2, O_2, O \tag{3.40}$$

Together with the direct production from ozone photolysis in the Hartley band, the concentration of $O_2(a^1\Delta_g)$ under photochemical steady state assumption can be derived as

$$[O_2(a^1\Delta_g)] = \frac{g_{IRA}[O_2] + \epsilon Q_2[O_2(b^1\Sigma_g^+)] + \phi_H J_H[O_3]}{A_3 + Q_3}, \tag{3.41}$$

where

$$Q_3 = k_3^{O_2}[O_2] + k_3^{N_2}[N_2] + k_3^O[O]. \tag{3.42}$$

In a photochemical model such as described above, the concentrations of $O(^1D)$, $O_2(b^1\Sigma_g^+)$ and $O_2(a^1\Delta_g)$ can be calculated, given the prescribed number density profiles of O_2, N_2 and O_3, as well as the temperature profile for some rate coefficients that have dependency on the background temperature. Figure 3.6 displays the results of the photochemical model in the altitude range of 60 km to 130 km, calculated for

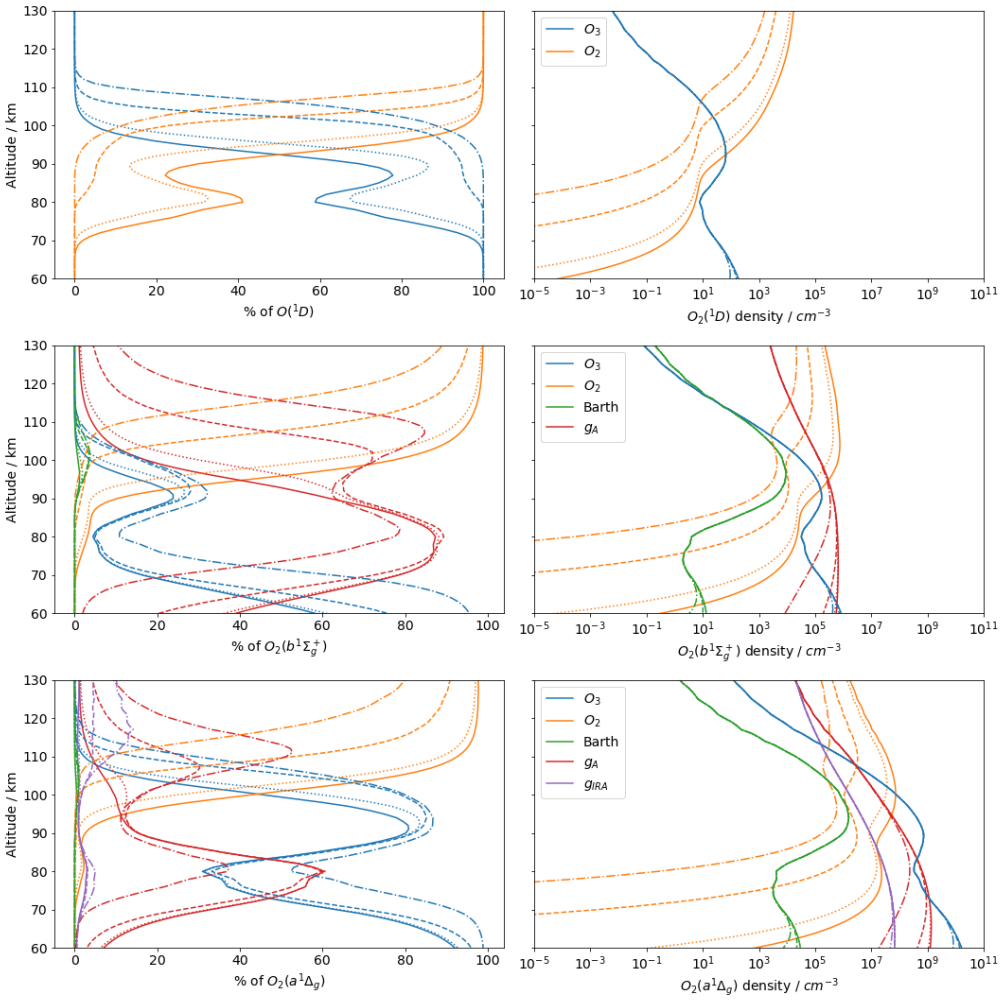

Figure 3.6: Altitude profiles of contributions from different sources (represented by colours, see text) to the total productions of $O(^1D)$ (first row), $O_2(b^1\Sigma_g^+)$ (second row) and $O_2(a^1\Delta_g)$(third row), while panels on the left show the relative contributions in percentage and panels on the right show the absolute contributions in number density of the species. Note that the horizontal axes of the right column are in logarithmic scale. Different line styles correspond to solar zenith angles of 30° (solid lines), 60° (dotted lines), 85° (dashed lines) and 90° (dash-dotted lines). Background density profiles of N_2, O_2 and temperature profile are taken from MSIS90 and O_3 taken from CMAM climatology in January at 15° N latitude.

four different solar zenith angle (SZA) in order to illustrate that the effect from the height of the sun varies between sources and in altitude. The interpretation on the significance of SZA also depends on whether is based on percentage-wise or absolute value. The label 'O_3' in Fig. 3.6 indicates productions through the ozone photolysis in the Hartley band, 'O_2' the molecular oxygen photolysis in the Schumann-Runge Continuum and Lyman α line, 'Barth' the three-body re-combination mechanism and 'g_A' and 'g_{IRA}' the solar excitation of the oxygen in the A-band and in the IRA-band, respectively. It can be seen that for the production of $O(^1D)$, ozone photolysis is more important below 90 km and oxygen photolysis above 90 km. For the productions of $O_2(b^1\Sigma_g^+)$ and $O_2(a^1\Delta_g)$, the general picture remains similar to $O(^1D)$, with the additional important sources from the solar excitation from the ground state O_2. The Barth-type mechanism contributes to the excited oxygen molecules to a lesser extend in comparison to other sources.

To conclude this section, we described the general kinetic scheme of the oxygen airglow photochemistry. Variations can be made by adding vibrational sub-levels or having different assumptions on the fractional efficiencies of each path. Nevertheless, it can be shown in all variations that ozone photolysis in the Hartley band is one of the most significant contributors to the production of both $O_2(b^1\Sigma_g^+)$ and $O_2(a^1\Delta_g)$ in their (secondary) peak region in the mesosphere. For this reason, these two airglow emissions can be used as proxies to infer the ozone density in the mesosphere during daytime, which is difficult to measure by other types of remote sensing technique.

Chapter 4

The inversion problem

4.1 Introduction

Remote sensing is one of the most effective ways to collect data over large areas, allowing us to study the global atmosphere. Specifically, compared to in-situ measurements in the mesosphere, typically by rocket sondes, sensing the region of interest from a spacecraft is relatively cost effective in a global perspective. Also, space-borne instruments are convenient for capturing the small variability in the middle and upper atmosphere, since the majority of the air mass is concentrated in the lower atmosphere that will otherwise dominate the signal. However, the quantity actually being measured is usually an indirect parameter which somehow is linked to the desired geophysical property. A typical example in atmospheric remote sensing is that the measured quantity is an electromagnetic signal, while the desired property is actually the temperature or the density of the air. Thus, an inverse problem arises when we attempt to find the best representation of the actual desired parameter.

Another example of the inverse problem arises from the viewing geometry when we attempt to study the vertical structure of various properties of the atmosphere. Such atmospheric sounding can be made by looking up to the sky from the ground, looking down to the ground from space (i.e., nadir sounding) or pointing to the side of Earth through a vertical range (i.e., limb sounding). All of the aforementioned sounding methods collect the integrated signal along the line-of-sight (LOS). However, the desired quantity is often the local signal at a given position, as if we would have sent a rocket sonde to the high atmosphere. In this thesis, limb sounding is the viewing geometry in focus, as we are sensing the mesosphere with the satellite Odin now and with MATS in the future.

The integrated electromagnetic signal that actually reaches the instrument is conventionally called the *limb radiance*, and the desired quantity to retrieve could for example be the volume emission rate (see Fig. 4.1 for an illustration). In such case, the inversion problem is how to 'unravel' the local volume emission given the total integrated limb radiance. Additionally, in Paper 1, we also discuss a different inversion problem which is to estimate the ozone density from the airglow volume emission rate.

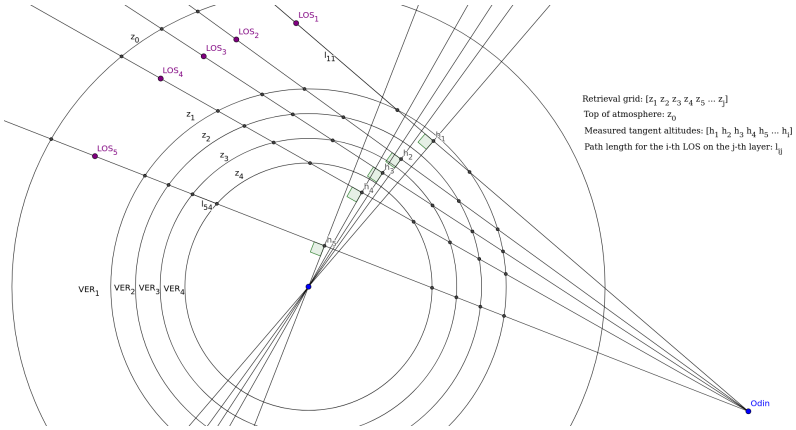

Figure 4.1: A graphical illustration of the limb geometry. VER denotes the volume emission rate, and LOS is line-of-sight.

In this chapter we will address a model based statistical approach to solve our inverse problems (rather than data driven methods like machine learning). The theory of how we can find the best estimates is based on a mathematical concept, that can be applied on different types of inversion problems. Although different mathematical approaches can be used to estimate the parameters, the physical relationship between the measured and desired quantities must be properly prescribed. The relationships we will consider in this thesis include: radiative transfer physics (from limb radiance to retrieve volume emission rate of IRA-band emissions) and the airglow photochemistry (from IRA-band emissions to retrieve ozone). However, we will not discuss these physical relationships further in this chapter. Instead, this chapter aims to provide 1) the basic mathematical foundation on inversion theory, more specifically the optimal estimation method and the Levenberg-Marquardt method, and 2) some highlights of the practical implementation in our application.

4.2 Theory of optimal estimation in Bayesian philosophy

Let us begin by defining the *measurement vector* $\mathbf{y}$ and the *state vector* $\mathbf{x}$, which correspond to the measured and the desired quantity that we attempt to estimate, such as limb radiance and ozone number density, respectively. The aforementioned physical relationship, that maps from the the state vector to the measurement vector, is the *forward model* $\mathbf{F}$, and the relationship can be expressed as

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon, \tag{4.1}$$

where ϵ is the random experimental error, e.g., that may come from the instrument's noise, which cannot be precisely described by $\mathbf{F}$.

Even though we do not know the exact value of ϵ (otherwise we could have corrected it in $\mathbf{y}$), we can assume the error term follows a certain probability distribution. The choice of the probability density function (pdf) depends on the

system that we are modelling, but usually it is assumed to be a Gaussian distribution (i.e., normal distribution), and so we will do in the rest of the thesis. The assumption of Gaussian distribution is justified by the need of a mathematically tractable model and being consistent with most problems in nature. In such a statistical approach, the unknown variable $\mathbf{x}$ that we attempt to estimate shall also be viewed as a pdf, i.e., as a realisation of a random variable, instead of a deterministic constant. In other words, the representation of the optimal solution of an inversion shall tell us the most likely $\mathbf{x}$, $\hat{\mathbf{x}}$, with an uncertainty characterised by a pdf. This is called the statistical approach.

Furthermore, some prior information about the state $\mathbf{x}_a$ can help to regularise the solution to be physically reasonable. This particularly applies to an ill-pose problem, which often is the case. Such a condition occurs when we attempt to retrieve more parameters than we actually can based on the amount of information given by the measurement that leads to multiple solutions, or having too much information that leads to inconsistency (no solution), or the uncertainty of the given information is too large that leads to a solution sensitive to noise (e.g., over fitting). This prior knowledge that we have before the actual measurement is made is termed the a priori (means 'from the earlier' in Latin). As we are following a statistical approach, the a priori knowledge shall also be described in terms of a pdf. Bayes' theorem helps us to mathematically combine the a priori pdf and the measurement pdf to find the desired a posteriori pdf (means 'from the later' in Latin). In other words, the process of inversion can be seen as updating the a priori information with the given measurements and uncertainties mapped to the state space. This is called the Bayesian approach since its implementation is based on Bayes' theorem.

A pdf of the Gaussian distribution is associated with two parameters: the mean (or expectated value) μ, and the standard deviation σ. The mean, as we can tell from its name, is the random variable most likely to be. The standard deviation is expressed in the form of covariance matrix which has elements of $S_{ij} = \rho_{ij}\sigma_i\sigma_j$ where the correlation coefficient ρ_{ij} has value of between -1 and 1. In this chapter, we use $\mathbf{S}_y$, $\mathbf{S}_x$ and $\mathbf{S}_a$ to denote the covariance matrix of the measurement vector, the estimated state vector and the a priori state vector, respectively.

In the subsequent sections, two inversion schemes that are implemented in this thesis will be presented. One applies to a linear forward model and another one to a non-linear forward model. The formalisation of the solutions are adopted from Rodgers (2000).

4.2.1 Linear optimal estimation

If the measurement vector $\mathbf{y}$ is linearly related to the state vector $\mathbf{x}$, the forward model $\mathbf{F}$ can be expressed in a linear form

$$\begin{aligned} \mathbf{y} &= \mathbf{F}(\mathbf{x}) + \epsilon \\ &= \mathbf{K}\mathbf{x} + \epsilon, \end{aligned} \tag{4.2}$$

where $\mathbf{K}$ is called the weighting function, or the Jacobian, which is a matrix comprised of all the first order partial derivatives of $\mathbf{F}(\mathbf{x})$ upon each element of $\mathbf{x}$, i.e., $\mathbf{K} =$

$\frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$.

If we neglect the error term, an immediate solution of a pure linear inversion problem, may seem to be

$$\mathbf{x} = \mathbf{K}^{-1}\mathbf{y} = \mathbf{G}\mathbf{y} \tag{4.3}$$

where $\mathbf{G}$ is the inverse matrix of $\mathbf{K}$, which compiles all the partial derivatives of the estimated state vector $\hat{\mathbf{x}}$ upon each element of the measurement vector $\mathbf{y}$, i.e., $\mathbf{G} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}}$. However, as we can expect, this simple inverse of the matrix $\mathbf{K}$ may be impossible to compute due to e.g., rank deficiency, thus the implementation of it is not so much practical in solving many physical problems. But this simple example is used to illustrate the general idea of the matrix $\mathbf{G}$ as a generalised inverse of $\mathbf{K}$, as later we will expand its formulation to more complex formats. $\mathbf{G}$ is called the *gain matrix* or *contribution function* in atmospheric remote sensing literature.

As mentioned earlier, the statistical approach is to make use of the full pdf of the measurement, which includes both the mean $\mathbf{y}$ and the covariance of the measurement $\mathbf{S}_y$, such that the cost function

$$\chi^2 = (\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{S}_y (\mathbf{y} - \mathbf{K}\mathbf{x}) \tag{4.4}$$

is minimised. The gain matrix and the best representation of the true state $\hat{\mathbf{x}}$ then become (detailed algebraic derivation is referred to Kay (1993))

$$\begin{aligned} \mathbf{G} &= (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1} \\ \hat{\mathbf{x}} &= \mathbf{G}\mathbf{y}. \end{aligned} \tag{4.5}$$

This solution is named the *weighted least square* (WLS) method, which allows the influence level to be different for each element in $\mathbf{y}$ to the estimate of the state vector, according to the uncertainty level. If the measurement noise is uniform and uncorrelated, then this method becomes the ordinary least square method (OLS, or LS).

However, as we want to incorporate the prior knowledge to regularise the retrieved quantity and prevent cases like over fitting, the pdf of the a priori, which includes both $\mathbf{x}_a$ and $\mathbf{S}_a$, shall also be utilised in the contribution function. In such case, the inversion can be viewed as an update from the a priori. Hence, the gain matrix $\mathbf{G}$, in this case, maps the difference between the $\mathbf{y}$ vector and the modelled a priori in measurement space to $(\hat{\mathbf{x}} - \mathbf{x}_a)$. The equations become

$$\begin{aligned} \mathbf{G} &= (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1} \\ \hat{\mathbf{x}} - \mathbf{x}_a &= \mathbf{G}(\mathbf{y} - \mathbf{K}\mathbf{x}_a) \end{aligned} \tag{4.6}$$

alternatively

$$\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{G}(\mathbf{y} - \mathbf{K}\mathbf{x}_a) \tag{4.7}$$

This solution is known as the optimal estimation method (OEM) or maximum a posteriori method (MAP). For this method, the procedure minimises the cost consisting of both the measurement noise and the a priori assumption, i.e.,

$$\chi^2 = (\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{S}_y (\mathbf{y} - \mathbf{K}\mathbf{x}) + (\hat{\mathbf{x}} - \mathbf{x}_a)^T \mathbf{S}_a (\hat{\mathbf{x}} - \mathbf{x}_a). \tag{4.8}$$

Now we have derived the best representation of the true state $\hat{\mathbf{x}}$, which is equivalent to the mean in Gaussian function. How about the standard deviation, i.e., the uncertainty, of the estimate? Thanks to the mathematical tractability property of the Gaussian pdf, we can evaluate the covariance of the retrieval uncertainty based on the measurement noise, (termed *retrieval noise* in Rodgers (2000)) following the form

$$\mathbf{S}_m = \mathbf{G}\mathbf{S}_y\mathbf{G}^T. \tag{4.9}$$

However, $\mathbf{S}_m$ is not the only component that contributes to the total error of the inversion. If OEM/MAP is used, another important source comes from the error in the a priori uncertainty, called the *smoothing error* evaluated by

$$\mathbf{S}_s = (\mathbf{A} - \mathbf{I})\mathbf{S}_a(\mathbf{A} - \mathbf{I})^T, \tag{4.10}$$

where $\mathbf{I}$ is the identity matrix which has the same size as $\mathbf{S}_a$, and $\mathbf{A}$ is named the averaging kernel,

$$\mathbf{A} = \mathbf{G}\mathbf{K} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}}. \tag{4.11}$$

It turns out that this $\mathbf{A}$ matrix gives us useful insights of the retrieval process, which tells us how sensitive the estimation is to the true state. In an ideal case, $\mathbf{A}$ would be an identity matrix $\mathbf{I}$ with the same size. This means that in each element of the estimated state, the changes will reflect the changes in the corresponding element of the true state in a ratio of one-to-one. In practice, as we have modified the formulation of $\mathbf{G}$ not to be the exact inverse of $\mathbf{K}$, the variable of the diagonal elements in the averaging kernel may be 'spread' to the off-diagonal elements. In fact, the spreading aspect reflects the true retrieval resolution of the retrieval being lower than the grid spaces predefined by the state vector. Thus, the actual (e.g., spatial) resolution shall not be confused by the predefined grid cell size.

4.2.2 Non-linear optimal estimation

A non-linear problem is in fact the most common type of problems in practice. The general approach to find the solution of a non-linear function is by using an iterative method. An iterative approach means that an algorithm re-calculates the same function with an updated input over and over again until a certain point that the solution being found is satisfied within a certain threshold. In an iterative procedure, the concept of the *initial guess* and *convergence* should be defined. The initial guess is the input that is initially to be set before any action so that the computation process can start at some point. Convergence is generally defined as the condition when the changes of variables of interest, compared to the previous iteration, are small enough. Usually a threshold is set as the convergence criterion to judge when the iteration should stop.

In the case of a non-linear inversion problem, the general idea of finding the solution is to linearise the forward model at a given state condition $\mathbf{F}(\mathbf{x}_n)$, e.g., initially $\mathbf{F}(\mathbf{x}_0)$. Such that at each iteration n the Jacobian matrix can be evaluated

as (following a forward Euler method)

$$\mathbf{K}_n = \frac{\mathbf{F}(\mathbf{x}_n + \delta\mathbf{x}) - \mathbf{F}(\mathbf{x}_n)}{\delta\mathbf{x}}, \tag{4.12}$$

where $\delta\mathbf{x}$ is the finite small perturbation introduce to compute the approximation of the derivative.

Following the previous formulation of the MAP method, the solution found at each iteration is

$$\begin{aligned} \mathbf{G}_n &= (\mathbf{K}_n^T\mathbf{S}_y^{-1}\mathbf{K}_n + \mathbf{S}_a^{-1})^{-1}\mathbf{K}_n^T\mathbf{S}_y^{-1} \\ \mathbf{x}_{n+1} &= \mathbf{x}_a + \mathbf{G}_n[\mathbf{y} - \mathbf{F}(\mathbf{x}_n) + \mathbf{K}_n(\mathbf{x}_n - \mathbf{x}_a)]. \end{aligned} \tag{4.13}$$

This procedure is to minimise the cost function

$$\chi_n^2 = (\mathbf{y} - \mathbf{F}(\mathbf{x}_n))^T\mathbf{S}_y(\mathbf{y} - \mathbf{F}(\mathbf{x}_n)) + (\mathbf{x}_n - \mathbf{x}_a)^T\mathbf{S}_a(\mathbf{x}_n - \mathbf{x}_a). \tag{4.14}$$

When the iteration number $n \to \infty$, the solution found shall be $\mathbf{x}_n \to \hat{\mathbf{x}}$. This method is called the Gauss-Newton method. When convergence is reached, the error analysis and characterisation are essentially the same for the linear case based on the Jacobian and the gain matrix at the last iteration, i.e.,

$$\begin{aligned} \mathbf{S}_m &= \mathbf{G}_n\mathbf{S}_y\mathbf{G}_n^T, \\ \mathbf{S}_s &= (\mathbf{A}_n - \mathbf{I})\mathbf{S}_a(\mathbf{A}_n - \mathbf{I})^T, \\ \mathbf{A}_n &= \mathbf{G}_n\mathbf{K}_n. \end{aligned} \tag{4.15}$$

By rearranging the equations for Gauss-Newton method, the solution can be expressed by $\mathbf{x}_{n+1}$ as a departure from $\mathbf{x}_n$ rather than $\mathbf{x}_a$

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{x}_n + \mathbf{H}^{-1}[\mathbf{K}_n^T\mathbf{S}_y^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x}_n)) - \mathbf{S}_a^{-1}(\mathbf{x}_n - \mathbf{x}_a)], \\ \mathbf{H} &= \mathbf{S}_a^{-1} + \mathbf{K}_n^T\mathbf{S}_y^{-1}\mathbf{K}_n, \end{aligned} \tag{4.16}$$

where $\mathbf{H}$ is known as the Hessian and represents the second derivative of the cost function, with the assumption that the second derivative of the forward model is negligible. In some cases, the iteration may result in an increasing residual, i.e., $|\mathbf{y} - \mathbf{F}(\mathbf{x}_n)|$ bigger than the previous iteration. Thus, the Levenberg-Marquardt method can be used to optimise the step size so that the cost function is 'guaranteed' to be minimised. This is achieved by adding a scaling term in the Hessian

$$\mathbf{H} = \mathbf{S}_a^{-1} + \mathbf{K}_n^T\mathbf{S}_y^{-1}\mathbf{K}_n + \gamma\mathbf{I}, \tag{4.17}$$

where $\mathbf{I}$ is an identity matrix and can be replaced by a scaling matrix $\mathbf{D}$ in order to scale accordingly to the dimensions and magnitudes of the elements in the state vector. γ is a scaling factor so that when $\gamma \to 0$ the inversion is close to the Gauss-Newton method, and when $\gamma \to \infty$ the inversion is close to the gradient decent method with a step size close to 0. Such a formulation of $\mathbf{H}$ indicates that there must be a value of γ to be found so that the optimisation step is meaningful. Thus, a process to search for a sufficiently large value of γ is needed if the cost function is increasing.

To conclude this section, the statistical approach is useful in determining the desired parameters as pdf rather than a deterministic constant. The mathematical formulations of several well developed methods, including WLS, OEM, Gauss-Newton and Levenberg-Marquardt are described here to treat linear and non-linear inversion problems.

4.3 Practical implementation

In practice, there are numerous aspects that are needed to be treated with caution when we carry out the implementations of the statistical method presented in the previous section. In this section, we discuss some of the highlights of these considerations in the inversion problems included in this thesis. In particular, the retrieval problem deals with the vertical profile of chemical species number density which exhibit high gradients.

A priori assumption

If the retrieval is based on a high response from the information given by the measurement, the resulting estimate of the state vector shall be, to a negligible extent, influenced by an arbitrarily assumed a priori state $\mathbf{x}_a$. Typically, the a priori state is taken from a mean value of independently measured results or model outputs to approximate the prior knowledge about the state vector.

Once the expected value of the prior knowledge, $\mathbf{x}_a$, is defined, the approximation for its standard deviation σ_a shall also be made in an appropriate manner. In our application, the elements of the state vector are expected to have several orders of magnitude difference (e.g., the atmospheric number density follows approximately an exponential function in altitude). If $\mathbf{x}_a$ is taken from a model output which resembles the true state reasonably well, a convenient way of defining σ_a is to choose values that are relative to the mean. For instance, $\sigma_a(i) = rx_a(i)$ for the element i where r is a factor scalar to tighten or loosen the constrains on the a priori state. In other words, the diagonal elements of the a priori covariance matrix

$$\mathrm{diag}(\mathbf{S}_a) = (r\mathbf{x}_a)^2. \tag{4.18}$$

Another type of constraint is to tackle the correlation between adjacent levels for the a priori assumption. As it is true for most of the atmospheric variables, elements that are closely spaced are less likely to have large oscillation, as compared to elements that are located far away from each other. Thus, the off-diagonal elements of the a priori covariance matrix can be defined by an ad-hoc constraint, for example

$$S_a(i,j) = \sigma_a(i)\sigma_a(j)\exp(-|i-j|\frac{\delta z}{h}), \tag{4.19}$$

where i, j are the element indices, δz is the physical spacing represented by the retrieval grid and h is the length scale which has the same unit with δz. This process prevents large oscillation in the retrieved solutions and thus is often termed the smoothing process.

Relative averaging kernel

As we previously proposed that the covariance of the a priori can be represented by a scaling factor multiplies the mean state $\mathbf{x}_a$, the averaging kernel can also be

viewed in a relative term to the a priori state, $\mathbf{A}^{rel}$. This can be implemented by the transformation from the ordinary averaging kernel $\mathbf{A}$

$$A^{rel}(i,j) = A(i,j)\frac{x_a(j)}{x_a(i)}. \tag{4.20}$$

The relative averaging kernel is more convenient to interpret when the state vector is expressed in a ratio depart from the a priori state.

Iterative computation

When solving a non-linear inversion, the derivative of the forward model is evaluated at each iteration numerically to calculate the Jacobian matrix. In practice, the small perturbation $\delta\mathbf{x}$ in Eq. 4.12 needs to be determined wisely to avoid any numerical instability limited by the computation hardware. Especially when the state vector is expected to have exponential gradient, $\delta\mathbf{x}$ should not be a uniform vector in which perturbations on all elements of the state vector are in the same distance. Here, a fractional approach can resolve this issue by allowing the small perturbation scaled with the expected state vector, e.g., $\delta\mathbf{x} = 0.001\mathbf{x}_a$.

Another common issue is the occurrence of non-physical values in the solution being found, for instance, due to low signal to noise ratio. Particularly, when a negative concentration of the certain chemical species is computed in the optimal solution found at an iteration, this $\mathbf{x}_n$ will be used in the next iteration for evaluating $\mathbf{F}(\mathbf{x}_n)$ as well as $\mathbf{K}_n$. Clearly this will introduce instability throughout the iterating process, since the forward model is based on a physical relationship. The measure being taken is to suppress all of the negative values occurred to be zero or a very small positive value, and the iterative process can then continue. Other methods include transforming the desired parameters of the inversion in a logarithmic scale. The disadvantage is that the error analysis and characterisation of the retrieval will all become logarithmic functions which complicates the interpretation of the inversion result.

The iterative approach to solve non-linear functions is to find an approximate solution step-wise until the 'correct' solution is found. As the iteration goes on and on, at some point we must set a threshold to tell the algorithm that the solution being found is good enough, or in the worst case the iteration may not even have converged to a solution. There is no 'best method' on which condition should be set for the judgement, but the condition on which iteration is stopped shall be documented in order to be used for filtering in the later data analysis. Here are a few criteria that have been adopted in this thesis.

Small change in x

When the solution found at the last iteration only shows a small difference to the one found in the previous iteration, it is said to be converged, for example,

$$\max\left(\frac{|\mathbf{x}_n - \mathbf{x}_{n-1}|}{\mathbf{x}_n}\right) < 0.02. \tag{4.21}$$

The d^2 test

Similar to the criterion for testing the change in the $\mathbf{x}$ vector, but now scaled by its estimated error

$$\begin{aligned} \hat{\mathbf{S}} &= (\mathbf{S}_a^{-1} + \mathbf{K}_n^T \mathbf{S}_y^{-1} \mathbf{K}_n)^{-1} \\ d^2 &= (\mathbf{x}_n - \mathbf{x}_{n-1})^T \hat{\mathbf{S}} (\mathbf{x}_n - \mathbf{x}_{n-1}). \end{aligned} \tag{4.22}$$

The value of d^2 can have a threshold of, e.g., 0.5 multiplied with the number of elements in $\mathbf{x}$, to judge convergence.

Lack of convergence

For the Levenberg-Marquardt method, the scaling parameter γ is set to ensure a reduction of the cost function, rather than an increase. In the searching process of γ, as we will call it the sub-iteration here, the following strategy is recommended in Rodgers (2000) and implemented in the algorithm for the ozone retrieval discussed in Paper 1:

- Start γ from an initial value of 10
- If the cost is smaller than the previous iteration, update $\mathbf{x}_n$ and multiply γ by a factor of 0.1 for the next iteration
- If the cost is bigger than the previous iteration, increase γ by a factor of 10 and try again

If γ infinitely increases to a large value but the cost is still not reducing, it might indicate that, e.g., there is a numerical problem, or the cost function has no minimum or the problem is completely under-constrained. Thus the iteration should be stopped.

To summarise this chapter, the mathematical formulations of several inversion methods including WLS, OEM/MAP, Gauss-Newton, Levenberg-Marquardt and their interpretations are discussed. The essence of the Bayesian statistical approach is to incorporate prior knowledge about the state for regularisation. In practical implementation of OEM/MAP (for linear problems) and Levenberg-Marquardt methods (for non-linear problems), there are several important considerations for the retrieval of atmospheric constituents' concentrations which roughly follow an exponential function in altitude.

Chapter 5

Summary of appended publications

Paper 1

Odin-OSIRIS consists of two optically independent instruments, the optical spectralgraph (OS) and the infrared imager (IRI). The latter routinely collects data of the oxygen infrared atmospheric band (IRA-band) emissions by one of its channels. The measurement principle and the geometry are very similar to the future mission MATS, thus the exploration of this IRI dataset certainly contributes to the preparatory process of the MATS mission which includes testing the retrieval algorithms, understanding the photochemical processes and possibly finding small scale dynamical features in the mesosphere.

In Paper 1 we presented a retrieval scheme to derive the mesospheric daytime ozone profiles from the IRI limb measurements of the IRA-band emissions (1.27 µm). The updated calibration scheme of the limb radiance data product is briefly described. Then, the optimal estimation method (OEM) is applied in the two inversion problems, the retrievals of volume emission rate (VER) of the $O_2(a^1\Delta_g)$ and the ozone number density profiles. For the VER retrieval, a linear forward model is used, which neglects self-absorption of the IRA-band along the line-of-sight in the radiative transfer modelling. For the O_3 retrieval, the forward model is non-linear due to the fact that the photochemistry model includes coefficients that depend on ozone concentrations. Here, the iterative Levenberg-Marquardt method is applied to derive O_3 concentrations from VER of oxygen IRA band emissions.

The high performance of the retrieval scheme is illustrated by the consistency between the IRI ozone product and two other ozone products obtained from the same spacecraft Odin, namely SMR and OS, even though their measurement principles are intrinsically different. It is also shown that the IRI ozone data has the advantage of high along-track resolution. The zonal mean monthly average profiles of IRI ozone are also compared with other independent ozone datasets onboard external satellites, namely ACE-FTS and MIPAS. These results indicate that the retrieval technique can be applied to process all the data collected throughout the 19 years of Odin

mission, leading to a long term mesospheric ozone dataset that can be used to study mesospheric dynamics and photochemistry.

Paper 2

The study was inspired by a previous publication by Dörnbrack et al. (2017) where they provide clear examples of how horizontal mean flow can affect the gravity wave signature to appear in ground-based lidar observation. Based on this publication we implemented the Doppler-shift into the gravity wave model for the preparation of the MATS mission (Li 2017). However, we failed to reproduce several important features of the wave-mean interaction (e.g. critical levels). We brought this question to Patrick Espy and his research group in Trondheim. We also brought the discussions directly with Andreas Dörnbrack and his research group in Oberpfaffenhofen. The results from the discussions and the main findings were then presented in 2019 at the EGU conference and serve as the basis for this manuscript.

In Paper 2, an alternative perspective on the interpretation of a ground-based lidar observation of a middle atmospheric gravity wave is discussed. This lidar observation was readily shown in a previous publication by Dörnbrack et al. (2017) where it was suggested that the recorded discontinuous pattern was possibly due to a wave packet superimposed on another wave in the background. However, in Paper 2 we provide evidences to show that the discontinuous pattern recorded can be explained by the quasi-monochromatic wave under an influence of a sudden change in the background wind shear. This interpretation is supported by the foundation of linear wave theory, as well as the investigation on the hourly-mean wind velocity taken from a combined product of the ECMWF reanalysis data (for lower altitudes) and the SLICE MF-radar (for higher altitudes) within the time frame and the location that the lidar observation was made.

In addition, this paper also attempts to bring clarity in some common nomenclature of different frequencies, that are specifically related to the Doppler-shift effect. We want to put attention on the analysis of the Doppler relationship, that the ground-based frequency must be the frequency observed in the same reference frame where the background wind velocity is recorded. Under the steady background condition, only the intrinsic frequency of a wave will be Doppler-shifted by the wind, rather than the ground-based frequency which is more closely linked to the source frequency. This concept of intrinsic frequency having the wind dependency may seem counter-intuitive as the word ‘intrinsic’ may refer to a constant property.

To conclude this study, the alternative interpretation suggests that the recorded signature is caused by the wave-mean, instead of the wave-wave, interaction. The simple model based on linear wave theory leads us to estimate the quasi-monochromatic wave with a phase-speed around $60\,\mathrm{ms}^{-1}$ and a travelling direction of south-south-east or north-north-west.

Chapter 6

Outlook

In this chapter, several possible research ideas will be discussed. They include the expansion of the current retrieval methods and atmospheric chemistry and dynamics studies by making use of the readily retrieved datasets.

6.1 Expansion of the retrievals

Tomographic method

As it has been demonstrated that the OSIRIS-IRI imager has the advantage of high sampling rate, a two dimensional tomographic retrieval technique can be applied to refine the resolution in along-track horizontal direction one step further. The main difference to the previous 1D inversion is that, a sequence of image exposures are used together for one inversion to retrieve a 2D atmospheric field, opposing to the retrieval of image by image. For example, the measurement vector $\mathbf{y}$ contains a number of pixels per image and b number of images, and the atmospheric state vector $\mathbf{x}$ contains volume emission rate (VER) represented in a 2D grid space at the orbital plane, with n and m numbers of grid points in vertical and horizontal directions, respectively. Thus the inversion problem to pose can be described by the following formulation

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1a} \\ \vdots \\ y_{b1} \\ \vdots \\ y_{ba} \end{bmatrix} = \mathbf{F}\left(\begin{bmatrix} x_{11} \\ \vdots \\ x_{1n} \\ \vdots \\ x_{m1} \\ \vdots \\ x_{mn} \end{bmatrix} \right) + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1a} \\ \vdots \\ \epsilon_{b1} \\ \vdots \\ \epsilon_{ba} \end{bmatrix} \tag{6.1}$$

where $\mathbf{F}$ is the forward model that transforms the VER to the measurements by all pixels considered.

The advantage of tomographic technique is to correctly allocate the origin of the atmospheric signal in along-track direction. However, there are several difficulties of the tomographic retrieval, one of which is being computationally costly. As we

can tell from the above equation, when the forward model is linearised, the Jacobian matrix $\mathbf{K}$ will contain $ab \times mn$ number of elements, and the covariance matrices $\mathbf{S}_a$ and $\mathbf{S}_y$ will have dimensions of $mn \times mn$ and $ab \times ab$, respectively. Even though the non-zero elements of the matrices can be stored wisely, such as in the form of sparse matrices, the computation is still substantially increased as compared to a 1D inversion.

Another difficulty for the tomographic retrieval method, perhaps not so trivial, comes from the scanning schedule of the Odin satellite compromised by the needs of the other instruments onboard. Shown as an example orbit in Paper 1 Fig. 3, the so-called mesospheric scanning mode creates data gaps up to 90 km penetrating through the important airglow layer. These data gaps eliminate many of the images in the orbit that can be useful for the tomographic retrieval. There are a substantial number of orbits that have the mesospheric scanning mode in the day part. Thus special treatments are needed for these orbits that have data gaps due to the scanning schedule.

Otherwise, IRI data that are obtained from the normal scanning mode and, even better, the staring mode can be readily used for the implementation of tomographic retrieval of the VERs. The result is expected to be beneficial for studying the small scale structures that IRI is capable to observe.

Inclusion of the IRA-band absorption

In the current retrieval scheme of the VER of $O_2(a^1\Delta_g)$, absorption by O_2 itself is not considered in the radiative transfer forward model so that the inversion is simplified. The assumption of a negligible absorption is valid for measurements that are taken at higher than ca. 60 km line-of-sight tangent where the air is optically thin. Thus, the retrieval only considered the IRI measurement pixels above this limit.

If the absorption is taken into account in the forward model, the retrieved VER and consequently the ozone profiles can reach down to lower altitudes. We will investigate methods to include this absorption in a simplified but accurate manner, so that the retrieval can reach below 60 km where the limb radiance is taken.

Inclusion of the IRA-band nighttime VER

Currently the retrieved VER of $O_2(a^1\Delta_g)$ is only available for daytime measurements, where the solar zenith angle is lower than 90°. The reason for this is that the linear retrieval scheme of OEM relies on the a priori VER profile which is calculated by the photochemistry model based on an ozone profile from a climatology (from CMAM).

At nighttime, most of the production mechanisms of $O_2(a^1\Delta_g)$ are 'switched off' and only the Barth-type mechanism is allowed in the photochemistry model. In this case, an atomic oxygen profile is needed to calculated the $O_2(a^1\Delta_g)$ density. In twilight conditions, the transition between day and night will require that a little more thought is put into the implementation of the photochemistry model, instead of a simple 'switch'. Due to the relatively long photochemical lifetime of $O_2(a^1\Delta_g)$ (about an hour), a time dependent model will be required.

Once the implementation of the time dependent photochemistry model is made, the retrieval of nighttime VER will be straight forward using the readily implemented retrieval scheme (OEM) and the available radiance data. The atomic oxygen profiles can then be retrieved in both day and night part of the orbit.

6.2 Mesospheric science studies

Assess $O_2(a^1\Delta_g)$ lifetime

Many reaction rate coefficients that are included in the photochemistry model described in Sect. 3.3 rely on laboratory experiments. One of these is the radiative lifetime of $O_2(a^1\Delta_g)$, i.e., the Einstein A coefficient, whose uncertainty is relatively large (e.g., G. Mlynczak and Olander (1995) has indicated a value of $1.47 \times 10^{-4}\,s^{-1}$ while the value of $2.58 \times 10^{-4}\,s^{-1}$ has been used previously, with a factor of 1.75 difference). This is mainly due to the long lifetime (about an hour) where laboratory-based measurements are difficult to perform because the decay of its population may partially be due to the collision with walls in an experimental setup. However, the mesosphere can serve as a 'natural laboratory' for such a measurement on estimation of radiative lifetime. This has been attempted in the past using ground based observations (Pendleton et al. 1996)

An attempt to estimate the $O_2(a^1\Delta_g)$ lifetime can be done by exploring the evening twilight measurements collected by IRI imager. When sunlight is not available for photolysis, the only source to produce $O_2(a^1\Delta_g)$ is by the Barth-type mechanism which is responsible for only 0.1 % in daytime. The rest should be the contribution by the delayed spontaneous emission which is also quenching by N_2 and O_2. Each orbit that crosses the terminator around the evening twilight can help us make a simple estimation. The remaining issue is the validity of the necessary assumptions, e.g., the longitudinal homogeneity in $O_2(a^1\Delta_g)$ density, background air density, quenching rates etc.

Exploration of the OH channel of IRI

The IRI imager also measures the radiance of the OH airglow in the Meinel (3-1) band. Particularly for the nighttime measurements, the airglow layer appears to show finer structures as compared to the $O_2(a^1\Delta_g)$ airglow. The smudging effect on fine scale structure in $O_2(a^1\Delta_g)$ is mainly due the relatively long lifetime. The short-lived OH can be used as a baseline to assess the level of smudging in $O_2(a^1\Delta_g)$ signal, as they are continuously measuring the same atmospheric volume.

Furthermore, the fine structure observed at the OH channel can readily be used to assess the gravity wave activities in the mesosphere. The difficulty lies in the separation of the mean and the wave perturbation.

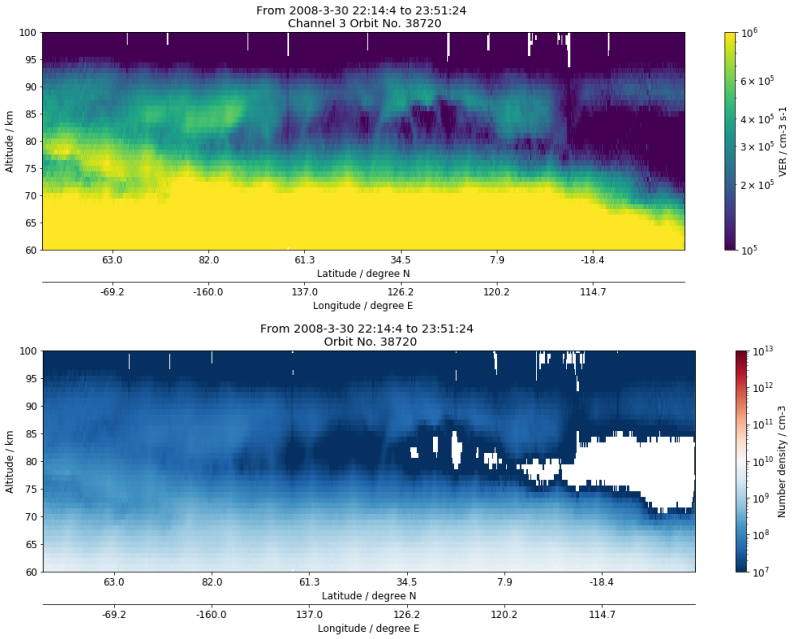

Figure 6.1: The volume emission rate of the oxygen IRA-band emission (top) and the ozone number density (bottom) measured in the day part of an example orbit (orbit No. 38720). This measurement is collected on March 20th, 2008. A large-scale feature appears from ca. 45° N (the beginning of the plot) to 82° N.

Tertiary ozone peak

The previously observed tertiary ozone peak at around 72 km is explained by the shortage of odd-hydrogen for the catalytic destruction of odd-oxygen, while the production of the odd-oxygen species is unchanged (Marsh, Smith, Guy Brasseur, et al. 2001). Unlike the primary and secondary ozone layers, the occurrence of the tertiary ozone maximum highly depends on season, latitude and altitude, according to multiple observations and modelling studies (e.g., Marsh, Smith, and Noble 2003; Kaufmann et al. 2003). Moreover, Degenstein et al. (2005) have demonstrated that IRI measurements of oxygen IRA-band and OH Meinel (3-1) band are capable of capturing the signature of the tertiary maximum. For a data visualisation of a two dimensional orbital scan of IRI shown in Fig. 6.1, this peak appears to be a finger-like, large-scale feature in the daytime portion of the orbit.

The daytime mesospheric ozone data product that is presented in Paper 1 can readily serve as a tool for further investigation on the behaviour of tertiary ozone peak. The main challenge is that the interpretation of data obtained from the Odin satellite requires careful considerations of orbital sampling pattern in latitude, solar zenith angle and local time. As Odin was launched in a dawn-dusk (6-18h) sun-synchronous orbit, the local time at the tangent point is rather constant but experiences a rapid change in the polar regions (e.g., from 6am to 6pm or vice versa). In addition, throughout the course of the mission, the orbit has on occasions drifted to be closer to 19h and 7h when crossing the ascending and descending node, respectively. Thus, when averaging profiles over a long time period, the local sampling pattern due to the satellite orbit must be carefully examined.

6.3 The future mission MATS

As the satellite MATS is scheduled to be launched in the end of 2020, numerous two dimensional images measuring the oxygen atmospheric band (emission from $O_2(b^1\Sigma_g{}^+)$) will be available for analysis. After the calibration process, the limb radiance can be inverted to VER using the retrieval scheme similar to the one described in Paper 1. Since the photochemical lifetime of $O_2(b^1\Sigma_g{}^+)$ is shorter than $O_2(a^1\Delta_g)$, finer structures due to small scale dynamics, such as gravity waves, are expected to be captured by MATS.

As the production of $O_2(b^1\Sigma_g{}^+)$ is related to the available ozone in the mesosphere under sunlit conditions, mesospheric ozone can be retrieved by using the same forward model as in Paper 1. During the nighttime it is the Barth-type mechanism that is responsible for the production of $O_2(b^1\Sigma_g{}^+)$, hence atomic oxygen density can be retrieved. In addition, the two spectral passbands in the infrared region will allow us to derive mesospheric temperature as the spectral filters are selected to be temperature sensitive. The retrieval scheme is similar to the one described in P. E. Sheese et al. (2010) where temperature is derived from the OSIRIS optical spectrograph measurement on the A-band emission.

Bibliography

Armstrong, EB (1982). "The association of visible airglow features with a gravity wave". In: *Journal of Atmospheric and Terrestrial Physics* 44.4, pp. 325–336 (cit. on p. 20).

Baumgarten, G., J. Fiedler, J. Hildebrand, and F. J. Lübken (2015). "Inertia gravity wave in the stratosphere and mesosphere observed by Doppler wind and temperature lidar". In: *Geophysical Research Letters*. ISSN: 19448007. DOI: 10.1002/2015GL066991 (cit. on p. 21).

Brasseur, G.P and Susan Solomon (2005). *Aeronomy of the Middle Atmosphere*. Vol. 53. 9, pp. 1689–1699. ISBN: 9788578110796. DOI: 10.1017/CBO9781107415324.004 (cit. on pp. 16, 29, 30).

Buhler, Oliver (2009). *Waves and Mean Flows*. Cambridge: Cambridge University Press. ISBN: 9780511605499. DOI: 10.1017/CBO9780511605499 (cit. on p. 13).

Clark, Terry L., Thomas Hauf, and Joachim P. Kuettner (1986). "Convectively forced internal gravity waves: Results from two-dimensional numerical experiments". In: *Quarterly Journal of the Royal Meteorological Society*. DOI: 10.1002/qj.49711247402 (cit. on p. 17).

Degenstein, Douglas Arthur, Richard L. Gattinger, Nick D. Lloyd, Adam E. Bourassa, Jonathan T. Wiensz, and Edam J. Llewellyn (Oct. 2005). "Observations of an extended mesospheric tertiary ozone peak". In: *Journal of Atmospheric and Solar-Terrestrial Physics* 67.15, pp. 1395–1402. ISSN: 13646826. DOI: 10.1016/j.jastp.2005.06.019 (cit. on p. 50).

Dörnbrack, Andreas, Sonja Gisinger, and Bernd Kaifler (2017). "On the interpretation of gravity wave measurements by ground-based lidars". In: *Atmosphere*. ISSN: 20734433. DOI: 10.3390/atmos8030049 (cit. on p. 46).

Ehard, B., B. Kaifler, N. Kaifler, and M. Rapp (2015). "Evaluation of methods for gravity wave extraction from middle-atmospheric lidar temperature measurements". In: *Atmospheric Measurement Techniques*. ISSN: 18678548. DOI: 10.5194/amt-8-4645-2015 (cit. on p. 18).

Fovell, Robert G., Dale Durran, and J.R. Holton (1992). "Numerical Simulations of Convectively Generated Stratospheric Gravity Waves". In: *Journal of the Atmospheric Sciences*. DOI: 10.1175/1520-0469(1992)049<1427:NSOCGS>2.0.CO;2 (cit. on p. 17).

Franzen, Christoph, Patrick Joseph Espy, Robert Edward Hibbins, and Anlaug Amanda Djupvik (Oct. 2018). "Observation of Quasiperiodic Structures in the Hydroxyl Airglow on Scales Below 100 m". In: *Journal of Geophysical Research: Atmospheres* 123.19, pp. 935–10. DOI: 10.1029/2018JD028732 (cit. on p. 19).

Fritts, David C. and M. Joan Alexander (2003). "Gravity wave dynamics and effects in the middle atmosphere". In: *Reviews of Geophysics*. DOI: 10.1029/2001RG000106 (cit. on pp. 11, 16, 18).

Gardner, Chester S and Michael J Taylor (1998). "Observational limits for lidar, radar, and airglow imager measurements of gravity wave parameters". In: *Journal of Geophysical Research: Atmospheres* 103.D6, pp. 6427–6437 (cit. on p. 19).

Gumbel, Jörg et al. (Jan. 2020). "The MATS satellite mission-Gravity wave studies by Mesospheric Airglow/Aerosol Tomography and Spectroscopy". In: *Atmospheric Chemistry and Physics* 20.1, pp. 431–455. DOI: 10.5194/acp-20-431-2020 (cit. on p. 5).

Heale, C. J. and J. B. Snively (2015). "Gravity wave propagation through a vertically and horizontally inhomogeneous background wind". In: *Journal of Geophysical Research*. ISSN: 21562202. DOI: 10.1002/2015JD023505 (cit. on pp. 13, 14).

Holton, James R. (1982). "The Role of Gravity Wave Induced Drag and Diffusion in the Momentum Budget of the Mesosphere". In: *Journal of the Atmospheric Sciences*. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1982)039<0791:TROGWI>2.0.CO;2 (cit. on pp. 9, 11, 16).

Kaifler, B., F. J. Lübken, J. Höffner, R. J. Morris, and T. P. Viehl (2015). "Lidar observations of gravity wave activity in the middle atmosphere over Davis (69°S, 78°E), Antarctica". In: *Journal of Geophysical Research*. ISSN: 21562202. DOI: 10.1002/2014JD022879 (cit. on p. 16).

Kaifler, N., B. Kaifler, B. Ehard, S. Gisinger, A. Dörnbrack, M. Rapp, R. Kivi, A. Kozlovsky, M. Lester, and B. Liley (2017). "Observational indications of downward-propagating gravity waves in middle atmosphere lidar data". In: *Journal of Atmospheric and Solar-Terrestrial Physics*. ISSN: 13646826. DOI: 10.1016/j.jastp.2017.03.003 (cit. on p. 16).

Karlsson, Bodil and Theodore G. Shepherd (June 2018). "The improbable clouds at the edge of the atmosphere". In: *Physics Today* 71.6, pp. 30–36. DOI: 10.1063/PT.3.3946 (cit. on p. 9).

Kaufmann, M., O. A. Gusev, K. U. Grossmann, F. J. Martín-Torres, D. R. Marsh, and A. A. Kutepov (May 2003). "Satellite observations of daytime and nighttime ozone in the mesosphere and lower thermosphere". In: *Journal of Geophysical Research D: Atmospheres* 108.9. DOI: 10.1029/2002jd002800 (cit. on p. 50).

Kay, Steven M. (1993). *Fundamentals of statistical signal processing.* xii, Prentice Hall PTR, 595 s. ISBN: 0133457117 (cit. on p. 38).

Koop, C. Gary (1981). "A Preliminary Investigation Of The Interaction Of Internal Gravity Waves With A Steady Shearing Motion". In: *Journal of Fluid Mechanics*. ISSN: 14697645. DOI: 10.1017/S0022112081003546 (cit. on pp. 15, 17).

Koop, C. Gary and Brian McGee (1986). "Measurements of internal gravity waves in a continuously stratified shear flow". In: *Journal of Fluid Mechanics*. ISSN: 14697645. DOI: 10.1017/S0022112086001817 (cit. on pp. 15, 17).

Krisch, Isabell et al. (2017). "First tomographic observations of gravity waves by the infrared limb imager GLORIA". In: *Atmospheric Chemistry and Physics*. ISSN: 16807324. DOI: 10.5194/acp-17-14937-2017 (cit. on p. 20).

Larsson, N, R Lilja, M Örth, S Söderholm, J Köhler, R Lindberg, and J Gumbel (2015). "InnoSat and MATS–An Ingenious Spacecraft Platform applied to Mesospheric Tomography and Spectroscopy". In: *10th IAA Symposium for Earth Observation* (cit. on p. 20).

Li, Anqi (2017). *A 3D-model for O2 airglow perturbations induced by gravity waves in the upper mesosphere (MSc Thesis)* (cit. on pp. 5, 7, 46).

Lighthill, M. J. (1967). "Waves in fluids". In: *Communications on Pure and Applied Mathematics*. ISSN: 10970312. DOI: 10.1002/cpa.3160200204 (cit. on p. 13).

Lindzen, R. S. (1981). "Turbulence and stress owing to gravity wave and tidal breakdown". In: *Journal of Geophysical Research* 86.C10, p. 9707. ISSN: 0148-0227. DOI: 10.1029/jc086ic10p09707 (cit. on p. 16).

Marks, Crispin J. and Stephen D. Eckermann (1995). "A Three-Dimensional Nonhydrostatic Ray-Tracing Model for Gravity Waves: Formulation and Preliminary Results for the Middle Atmosphere". In: *Journal of the Atmospheric Sciences*. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1995)052<1959:ATDNRT>2.0.CO;2 (cit. on p. 13).

Marsh, Daniel, Anne Smith, Guy Brasseur, Martin Kaufmann, and Klaus Grossmann (Dec. 2001). "The existence of a tertiary ozone maximum in the high-latitude middle mesosphere". In: *Geophysical Research Letters* 28.24, pp. 4531–4534. ISSN: 00948276. DOI: 10.1029/2001GL013791 (cit. on p. 50).

Marsh, Daniel, Anne Smith, and Erik Noble (Feb. 2003). "Mesospheric ozone response to changes in water vapor". In: *Journal of Geophysical Research D: Atmospheres* 108.3. DOI: 10.1029/2002jd002705 (cit. on p. 50).

McDade, I.C., D.P. Murtagh, R.G.H. Greer, P.H.G. Dickinson, G. Witt, J. Stegman, E.J. Llewellyn, L. Thomas, and D.B. Jenkins (Sept. 1986). "ETON 2: Quenching parameters for the proposed precursors of O2(b1Σg+) and O(1S) in the terrestrial nightglow". In: *Planetary and Space Science* 34.9, pp. 789–800. ISSN: 00320633. DOI: 10.1016/0032-0633(86)90075-9 (cit. on p. 31).

McLandress, Charles (1998). "On the importance of gravity waves in the middle atmosphere and their parameterization in general circulation models". In: *Journal of Atmospheric and Solar-Terrestrial Physics*. ISSN: 13646826. DOI: 10.1016/S1364-6826(98)00061-3 (cit. on p. 9).

Medeiros, A F, M J Taylor, H Takahashi, P P Batista, and D Gobbi (2003). "An investigation of gravity wave activity in the low-latitude upper mesosphere: Propagation direction and wind filtering". In: *J. Geophys. Res* 108.441110. DOI: 10.1029/2002JD002593 (cit. on p. 20).

Mlynczak, G and S Olander (1995). "On the utility of the molecular oxygen dayglow emissions as proxies for middle atmospheric ozone". In: *Geophysical Research Letters* 22, pp. 1377–1380 (cit. on p. 49).

Mlynczak, M. G., F. Morgan, J. H. Yee, P. Espy, D. Murtagh, B. Marshall, and F. Schmidlin (Mar. 2001). "Simultaneous measurements of the O2(1Δ) and O2(1Σ) airglows and ozone in the daytime mesosphere". In: *Geophysical Research Letters* 28.6, pp. 999–1002. ISSN: 00948276. DOI: 10.1029/2000GL012423 (cit. on p. 28).

Mlynczak, Martin G., B. Thomas Marshall, F. Javier Martin-Torres, James M. Russell, R. Earl Thompson, Ellis E. Remsberg, and Larry L. Gordley (Aug. 2007).

"Sounding of the Atmosphere using Broadband Emission Radiometry observations of daytime mesospheric O2 (1Δ) 1.27 μm emission and derivation of ozone, atomic oxygen, and solar and chemical energy deposition rates". In: *Journal of Geophysical Research Atmospheres* 112.15. DOI: 10.1029/2006JD008355 (cit. on p. 28).

Murtagh, D. P., G. Witt, J. Stegman, I. C. McDade, E. J. Llewellyn, F. Harris, and R. G H Greer (1990). "An assessment of proposed O(1S) and O2 (b 1Σg+)nightglow excitation parameters". In: *Planetary and Space Science* 38.1, pp. 43–53. ISSN: 00320633. DOI: 10.1016/0032-0633(90)90004-A (cit. on pp. 31, 32).

Murtagh, Donal et al. (2002). "An overview of the Odin atmospheric mission". In: *Canadian Journal of Physics* 80.4, pp. 309–319. DOI: 10.1139/P01-157 (cit. on p. 5).

Nakamura, T, A Higashikawa, T Tsuda, and Y Matsushita (1999). "Seasonal variations of gravity wave structures in OH airglow with a CCD imager at Shigaraki". In: *Earth, planets and space* 51.7-8, pp. 897–906 (cit. on p. 20).

Nappo, C J (2002). *An Introduction to Atmospheric Gravity Waves*. English. International Geophysics Series v. 85. San Diego: Academic Press. ISBN: 9780125140829 (cit. on p. 11).

Pendleton, W. R., D. J. Baker, R. J. Reese, and R. R. O'Neil (1996). "Decay of O2(a1Δg) in the evening twilight airglow: Implications for the radiative lifetime". In: *Geophysical Research Letters* 23.9, pp. 1013–1016. ISSN: 19448007. DOI: 10.1029/96GL00946 (cit. on p. 49).

Preusse, Peter, Andreas Dörnbrack, Stephen D. Eckermann, Martin Riese, Bernd Schaeler, Julio T. Bacmeister, Dave Broutman, and Klaus U. Grossmann (2002). "Space-based measurements of stratospheric mountain waves by CRISTA 1. Sensitivity, analysis method, and a case study". In: *Journal of Geophysical Research: Atmospheres*. ISSN: 21698996. DOI: 10.1029/2001JD000699 (cit. on p. 20).

Preusse, Peter, Stephen D. Eckermann, and Manfred Ern (2008). "Transparency of the atmosphere to short horizontal wavelength gravity waves". In: *Journal of Geophysical Research Atmospheres*. ISSN: 01480227. DOI: 10.1029/2007JD009682 (cit. on p. 19).

Prusa, Joseph M, Piotr K Smolarkiewicz, and Rolando R Garcia (1996). "Propagation and Breaking at High Altitudes of Gravity Waves Excited by Tropospheric Forcing". In: *Journal of the Atmospheric Sciences*. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1996)053<2186:pabaha>2.0.co;2 (cit. on p. 17).

Rodgers, Clive D (2000). *Inverse methods for atmospheric sounding : theory and practice*. Series on atmospheric, oceanic and planetary physics: 2. World Scientific. ISBN: 981022740X (cit. on pp. 37, 39, 43).

Sheese, P. E., E. J. Llewellyn, R. L. Gattinger, A. E. Bourassa, D. A. Degenstein, N. D. Lloyd, and I. C. McDade (2010). "Temperatures in the upper mesosphere and lower thermosphere from OSIRIS observations of O2 A-band emission spectra". In: *Canadian Journal of Physics* 88.12, pp. 919–925. DOI: 10.1139/p10-093 (cit. on p. 51).

Sheese, Patrick (Jan. 2009). "Mesospheric ozone densities retrieved from OSIRIS observations of the O2 A-band dayglow". PhD thesis (cit. on p. 28).

Song, Rui, Martin Kaufmann, Jörn Ungermann, Manfred Ern, Guang Liu, and Martin Riese (2017). "Tomographic reconstruction of atmospheric gravity wave parameters from airglow observations". In: *Atmos. Meas. Tech* 10, pp. 4601–4612. DOI: 10.5194/amt-10-4601-2017 (cit. on p. 20).

Thomas, R. J., C. A. Barth, D. W. Rusch, and R. W. Sanders (1984). "Solar Mesosphere Explorer near-infrared spectrometer: measurements of 1.27 micrometer radiances and the inference of mesospheric ozone." In: *Journal of Geophysical Research*. ISSN: 01480227. DOI: 10.1029/JD089iD06p09569 (cit. on p. 28).

Ungermann, Jörn (2011). *Tomographic reconstruction of atmospheric volumes from infrared limb-imager measurements*. Vol. 106. Forschungszentrum Jülich (cit. on p. 20).

Vadas, Sharon L. and Erich Becker (2018). "Numerical Modeling of the Excitation, Propagation, and Dissipation of Primary and Secondary Gravity Waves during Wintertime at McMurdo Station in the Antarctic". In: *Journal of Geophysical Research: Atmospheres*. ISSN: 21698996. DOI: 10.1029/2017JD027974 (cit. on pp. 16, 18).

Vadas, Sharon L., Jian Zhao, Xinzhao Chu, and Erich Becker (Sept. 2018). "The Excitation of Secondary Gravity Waves From Local Body Forces: Theory and Observation". In: *Journal of Geophysical Research: Atmospheres* 123.17, pp. 9296–9325. DOI: 10.1029/2017JD027970 (cit. on pp. 16, 18).

Wu, Dong L., Peter Preusse, Stephen D. Eckermann, Jonathan H. Jiang, Manuel de la Torre Juarez, Lawrence Coy, and Ding Y. Wang (2006). "Remote sounding of atmospheric gravity waves with satellite limb and nadir techniques". In: *Advances in Space Research*. ISSN: 02731177. DOI: 10.1016/j.asr.2005.07.031 (cit. on p. 19).

Yankovsky, V A and R O Manuilova (2006). "Model of daytime emissions of electronically-vibrationally excited products of O_3 and O_2 photolysis: application to ozone retrieval". In: *Annales Geophysicae* 24.11, pp. 2823–2839. DOI: 10.5194/angeo-24-2823-2006 (cit. on p. 28).

Yankovsky, Valentine A., Kseniia V. Martyshenko, Rada O. Manuilova, and Artem G. Feofilov (Sept. 2016). "Oxygen dayglow emissions as proxies for atomic oxygen and ozone in the mesosphere and lower thermosphere". In: *Journal of Molecular Spectroscopy* 327, pp. 209–231. ISSN: 1096083X. DOI: 10.1016/j.jms.2016.03.006 (cit. on p. 28).